

BIRCH, STEWART, KOLASCH & BIRCH, LLP

INTELLECTUAL PROPERTY LAW
8110 GATEHOUSE ROAD
SUITE 500 EAST
FALLS CHURCH, VA 22042-1210
U S A
(703) 205-8000

FAX: (703) 205-8050
(703) 698-8590 (GIV)

e-mail: mailroom@bskb.com
web: http://www.bskb.com

CALIFORNIA OFFICE:
COSTA MESA, CALIFORNIA

THOMAS S. AUCHTERLONIE
JAMES T. ELLER, JR.
SCOTT L. LEWIS
MARK J. NUELL, PH D
D. RICHARD ANDERSON
PAUL C. LEWIS
MARK W. MILSTEAD*
JOHN CAMPA
RICHARD J. GALLAGHER

REG. PATENT AGENTS:
FREDERICK R. HANDREN
MARYANNE ARMSTRONG, PH D
MAKI HATSUMI
MIKE S. RYU
CRAIG A. MCROBBIE
GARTH M. DAHLEN, PH D.
LAURA C. LUTZ
ROBERT E. GOZZNER, PH D
HYUNG N. SOHN
MATTHEW J. LATTIG
ALAN PEDERSEN GILES
JUSTIN D. KARJALA
C. KEITH MONTGOMERY
TIMOTHY R. WYCKOFF
HERMES M. SOYEZ, PH D
KRISTIN L. RUPERT, PH D

10486 U.S. PTO
09/595326
06/16/00

00/97/90
01/01/00
02/02/00
03/03/00
04/04/00
05/05/00
06/06/00
07/07/00
08/08/00
09/09/00
10/10/00
11/11/00
12/12/00
01/01/01
02/02/01
03/03/01
04/04/01
05/05/01
06/06/01
07/07/01
08/08/01
09/09/01
10/10/01
11/11/01
12/12/01
01/01/02
02/02/02
03/03/02
04/04/02
05/05/02
06/06/02
07/07/02
08/08/02
09/09/02
10/10/02
11/11/02
12/12/02
01/01/03
02/02/03
03/03/03
04/04/03
05/05/03
06/06/03
07/07/03
08/08/03
09/09/03
10/10/03
11/11/03
12/12/03
01/01/04
02/02/04
03/03/04
04/04/04
05/05/04
06/06/04
07/07/04
08/08/04
09/09/04
10/10/04
11/11/04
12/12/04
01/01/05
02/02/05
03/03/05
04/04/05
05/05/05
06/06/05
07/07/05
08/08/05
09/09/05
10/10/05
11/11/05
12/12/05
01/01/06
02/02/06
03/03/06
04/04/06
05/05/06
06/06/06
07/07/06
08/08/06
09/09/06
10/10/06
11/11/06
12/12/06
01/01/07
02/02/07
03/03/07
04/04/07
05/05/07
06/06/07
07/07/07
08/08/07
09/09/07
10/10/07
11/11/07
12/12/07
01/01/08
02/02/08
03/03/08
04/04/08
05/05/08
06/06/08
07/07/08
08/08/08
09/09/08
10/10/08
11/11/08
12/12/08
01/01/09
02/02/09
03/03/09
04/04/09
05/05/09
06/06/09
07/07/09
08/08/09
09/09/09
10/10/09
11/11/09
12/12/09
01/01/10
02/02/10
03/03/10
04/04/10
05/05/10
06/06/10
07/07/10
08/08/10
09/09/10
10/10/10
11/11/10
12/12/10
01/01/11
02/02/11
03/03/11
04/04/11
05/05/11
06/06/11
07/07/11
08/08/11
09/09/11
10/10/11
11/11/11
12/12/11
01/01/12
02/02/12
03/03/12
04/04/12
05/05/12
06/06/12
07/07/12
08/08/12
09/09/12
10/10/12
11/11/12
12/12/12
01/01/13
02/02/13
03/03/13
04/04/13
05/05/13
06/06/13
07/07/13
08/08/13
09/09/13
10/10/13
11/11/13
12/12/13
01/01/14
02/02/14
03/03/14
04/04/14
05/05/14
06/06/14
07/07/14
08/08/14
09/09/14
10/10/14
11/11/14
12/12/14
01/01/15
02/02/15
03/03/15
04/04/15
05/05/15
06/06/15
07/07/15
08/08/15
09/09/15
10/10/15
11/11/15
12/12/15
01/01/16
02/02/16
03/03/16
04/04/16
05/05/16
06/06/16
07/07/16
08/08/16
09/09/16
10/10/16
11/11/16
12/12/16
01/01/17
02/02/17
03/03/17
04/04/17
05/05/17
06/06/17
07/07/17
08/08/17
09/09/17
10/10/17
11/11/17
12/12/17
01/01/18
02/02/18
03/03/18
04/04/18
05/05/18
06/06/18
07/07/18
08/08/18
09/09/18
10/10/18
11/11/18
12/12/18
01/01/19
02/02/19
03/03/19
04/04/19
05/05/19
06/06/19
07/07/19
08/08/19
09/09/19
10/10/19
11/11/19
12/12/19
01/01/20
02/02/20
03/03/20
04/04/20
05/05/20
06/06/20
07/07/20
08/08/20
09/09/20
10/10/20
11/11/20
12/12/20
01/01/21
02/02/21
03/03/21
04/04/21
05/05/21
06/06/21
07/07/21
08/08/21
09/09/21
10/10/21
11/11/21
12/12/21
01/01/22
02/02/22
03/03/22
04/04/22
05/05/22
06/06/22
07/07/22
08/08/22
09/09/22
10/10/22
11/11/22
12/12/22
01/01/23
02/02/23
03/03/23
04/04/23
05/05/23
06/06/23
07/07/23
08/08/23
09/09/23
10/10/23
11/11/23
12/12/23
01/01/24
02/02/24
03/03/24
04/04/24
05/05/24
06/06/24
07/07/24
08/08/24
09/09/24
10/10/24
11/11/24
12/12/24
01/01/25
02/02/25
03/03/25
04/04/25
05/05/25
06/06/25
07/07/25
08/08/25
09/09/25
10/10/25
11/11/25
12/12/25
01/01/26
02/02/26
03/03/26
04/04/26
05/05/26
06/06/26
07/07/26
08/08/26
09/09/26
10/10/26
11/11/26
12/12/26
01/01/27
02/02/27
03/03/27
04/04/27
05/05/27
06/06/27
07/07/27
08/08/27
09/09/27
10/10/27
11/11/27
12/12/27
01/01/28
02/02/28
03/03/28
04/04/28
05/05/28
06/06/28
07/07/28
08/08/28
09/09/28
10/10/28
11/11/28
12/12/28
01/01/29
02/02/29
03/03/29
04/04/29
05/05/29
06/06/29
07/07/29
08/08/29
09/09/29
10/10/29
11/11/29
12/12/29
01/01/30
02/02/30
03/03/30
04/04/30
05/05/30
06/06/30
07/07/30
08/08/30
09/09/30
10/10/30
11/11/30
12/12/30
01/01/31
02/02/31
03/03/31
04/04/31
05/05/31
06/06/31
07/07/31
08/08/31
09/09/31
10/10/31
11/11/31
12/12/31
01/01/32
02/02/32
03/03/32
04/04/32
05/05/32
06/06/32
07/07/32
08/08/32
09/09/32
10/10/32
11/11/32
12/12/32
01/01/33
02/02/33
03/03/33
04/04/33
05/05/33
06/06/33
07/07/33
08/08/33
09/09/33
10/10/33
11/11/33
12/12/33
01/01/34
02/02/34
03/03/34
04/04/34
05/05/34
06/06/34
07/07/34
08/08/34
09/09/34
10/10/34
11/11/34
12/12/34
01/01/35
02/02/35
03/03/35
04/04/35
05/05/35
06/06/35
07/07/35
08/08/35
09/09/35
10/10/35
11/11/35
12/12/35
01/01/36
02/02/36
03/03/36
04/04/36
05/05/36
06/06/36
07/07/36
08/08/36
09/09/36
10/10/36
11/11/36
12/12/36
01/01/37
02/02/37
03/03/37
04/04/37
05/05/37
06/06/37
07/07/37
08/08/37
09/09/37
10/10/37
11/11/37
12/12/37
01/01/38
02/02/38
03/03/38
04/04/38
05/05/38
06/06/38
07/07/38
08/08/38
09/09/38
10/10/38
11/11/38
12/12/38
01/01/39
02/02/39
03/03/39
04/04/39
05/05/39
06/06/39
07/07/39
08/08/39
09/09/39
10/10/39
11/11/39
12/12/39
01/01/40
02/02/40
03/03/40
04/04/40
05/05/40
06/06/40
07/07/40
08/08/40
09/09/40
10/10/40
11/11/40
12/12/40
01/01/41
02/02/41
03/03/41
04/04/41
05/05/41
06/06/41
07/07/41
08/08/41
09/09/41
10/10/41
11/11/41
12/12/41
01/01/42
02/02/42
03/03/42
04/04/42
05/05/42
06/06/42
07/07/42
08/08/42
09/09/42
10/10/42
11/11/42
12/12/42
01/01/43
02/02/43
03/03/43
04/04/43
05/05/43
06/06/43
07/07/43
08/08/43
09/09/43
10/10/43
11/11/43
12/12/43
01/01/44
02/02/44
03/03/44
04/04/44
05/05/44
06/06/44
07/07/44
08/08/44
09/09/44
10/10/44
11/11/44
12/12/44
01/01/45
02/02/45
03/03/45
04/04/45
05/05/45
06/06/45
07/07/45
08/08/45
09/09/45
10/10/45
11/11/45
12/12/45
01/01/46
02/02/46
03/03/46
04/04/46
05/05/46
06/06/46
07/07/46
08/08/46
09/09/46
10/10/46
11/11/46
12/12/46
01/01/47
02/02/47
03/03/47
04/04/47
05/05/47
06/06/47
07/07/47
08/08/47
09/09/47
10/10/47
11/11/47
12/12/47
01/01/48
02/02/48
03/03/48
04/04/48
05/05/48
06/06/48
07/07/48
08/08/48
09/09/48
10/10/48
11/11/48
12/12/48
01/01/49
02/02/49
03/03/49
04/04/49
05/05/49
06/06/49
07/07/49
08/08/49
09/09/49
10/10/49
11/11/49
12/12/49
01/01/50
02/02/50
03/03/50
04/04/50
05/05/50
06/06/50
07/07/50
08/08/50
09/09/50
10/10/50
11/11/50
12/12/50
01/01/51
02/02/51
03/03/51
04/04/51
05/05/51
06/06/51
07/07/51
08/08/51
09/09/51
10/10/51
11/11/51
12/12/51
01/01/52
02/02/52
03/03/52
04/04/52
05/05/52
06/06/52
07/07/52
08/08/52
09/09/52
10/10/52
11/11/52
12/12/52
01/01/53
02/02/53
03/03/53
04/04/53
05/05/53
06/06/53
07/07/53
08/08/53
09/09/53
10/10/53
11/11/53
12/12/53
01/01/54
02/02/54
03/03/54
04/04/54
05/05/54
06/06/54
07/07/54
08/08/54
09/09/54
10/10/54
11/11/54
12/12/54
01/01/55
02/02/55
03/03/55
04/04/55
05/05/55
06/06/55
07/07/55
08/08/55
09/09/55
10/10/55
11/11/55
12/12/55
01/01/56
02/02/56
03/03/56
04/04/56
05/05/56
06/06/56
07/07/56
08/08/56
09/09/56
10/10/56
11/11/56
12/12/56
01/01/57
02/02/57
03/03/57
04/04/57
05/05/57
06/06/57
07/07/57
08/08/57
09/09/57
10/10/57
11/11/57
12/12/57
01/01/58
02/02/58
03/03/58
04/04/58
05/05/58
06/06/58
07/07/58
08/08/58
09/09/58
10/10/58
11/11/58
12/12/58
01/01/59
02/02/59
03/03/59
04/04/59
05/05/59
06/06/59
07/07/59
08/08/59
09/09/59
10/10/59
11/11/59
12/12/59
01/01/60
02/02/60
03/03/60
04/04/60
05/05/60
06/06/60
07/07/60
08/08/60
09/09/60
10/10/60
11/11/60
12/12/60
01/01/61
02/02/61
03/03/61
04/04/61
05/05/61
06/06/61
07/07/61
08/08/61
09/09/61
10/10/61
11/11/61
12/12/61
01/01/62
02/02/62
03/03/62
04/04/62
05/05/62
06/06/62
07/07/62
08/08/62
09/09/62
10/10/62
11/11/62
12/12/62
01/01/63
02/02/63
03/03/63
04/04/63
05/05/63
06/06/63
07/07/63
08/08/63
09/09/63
10/10/63
11/11/63
12/12/63
01/01/64
02/02/64
03/03/64
04/04/64
05/05/64
06/06/64
07/07/64
08/08/64
09/09/64
10/10/64
11/11/64
12/12/64
01/01/65
02/02/65
03/03/65
04/04/65
05/05/65
06/06/65
07/07/65
08/08/65
09/09/65
10/10/65
11/11/65
12/12/65
01/01/66
02/02/66
03/03/66
04/04/66
05/05/66
06/06/66
07/07/66
08/08/66
09/09/66
10/10/66
11/11/66
12/12/66
01/01/67
02/02/67
03/03/67
04/04/67
05/05/67
06/06/67
07/07/67
08/08/67
09/09/67
10/10/67
11/11/67
12/12/67
01/01/68
02/02/68
03/03/68
04/04/68
05/05/68
06/06/68
07/07/68
08/08/68
09/09/68
10/10/68
11/11/68
12/12/68
01/01/69
02/02/69
03/03/69
04/04/69
05/05/69
06/06/69
07/07/69
08/08/69
09/09/69
10/10/69
11/11/69
12/12/69
01/01/70
02/02/70
03/03/70
04/04/70
05/05/70
06/06/70
07/07/70
08/08/70
09/09/70
10/10/70
11/11/70
12/12/70
01/01/71
02/02/71
03/03/71
04/04/71
05/05/71
06/06/71
07/07/71
08/08/71
09/09/71
10/10/71
11/11/71
12/12/71
01/01/72
02/02/72
03/03/72
04/04/72
05/05/72
06/06/72
07/07/72
08/08/72
09/09/72
10/10/72
11/11/72
12/12/72
01/01/73
02/02/73
03/03/73
04/04/73
05/05/73
06/06/73
07/07/73
08/08/73
09/09/73
10/10/73
11/11/73
12/12/73
01/01/74
02/02/74
03/03/74
04/04/74
05/05/74
06/06/74
07/07/74
08/08/74
09/09/74
10/10/74
11/11/74
12/12/74
01/01/75
02/02/75
03/03/75
04/04/75
05/05/75
06/06/75
07/07/75
08/08/75
09/09/75
10/10/75
11/11/75
12/12/75
01/01/76
02/02/76
03/03/76
04/04/76
05/05/76
06/06/76
07/07/76
08/08/76
09/09/76
10/10/76
11/11/76
12/12/76
01/01/77
02/02/77
03/03/77
04/04/77
05/05/77
06/06/77
07/07/77
08/08/77
09/09/77
10/10/77
11/11/77
12/12/77
01/01/78
02/02/78
03/03/78
04/04/78
05/05/78
06/06/78
07/07/78
08/08/78
09/09/78
10/10/78
11/11/78
12/12/78
01/01/79
02/02/79
03/03/79
04/04/79
05/05/79
06/06/79
07/07/79
08/08/79
09/09/79
10/10/79
11/11/79
12/12/79
01/01/80
02/02/80
03/03/80
04/04/80
05/05/80
06/06/80
07/07/80
08/08/80
09/09/80
10/10/80
11/11/80
12/12/80
01/01/81
02/02/81
03/03/81
04/04/81
05/05/81
06/06/81
07/07/81
08/08/81
09/09/81
10/10/81
11/11/81
12/12/81
01/01/82
02/02/82
03/03/82
04/04/82
05/05/82
06/06/82
07/07/82
08/08/82
09/09/82
10/10/82
11/11/82
12/12/82
01/01/83
02/02/83
03/03/83
04/04/83
05/05/83
06/06/83
07/07/83
08/08/83
09/09/83
10/10/83
11/11/83
12/12/83
01/01/84
02/02/84
03/03/84
04/04/84
05/05/84
06/06/84
07/07/84
08/08/84
09/09/84
10/10/84
11/11/84
12/12/84
01/01/85
02/02/85
03/03/85
04/04/85
05/05/85
06/06/85
07/07/85
08/08/85
09/09/85
10/10/85
11/11/85
12/12/85
01/01/86
02/02/86
03/03/86
04/04/86
05/05/86
06/06/86
07/07/86
08/08/86
09/09/86
10/10/86
11/11/86
12/12/86
01/01/87
02/02/87
03/03/87
04/04/87
05/05/87
06/06/87
07/07/87
08/08/87
09/09/87
10/10/87
11/11/87
12/12/87
01/01/88
02/02/88
03/03/88
04/04/88
05/05/88
06/06/88
07/07/88
08/08/88
09/09/88
10/10/88
11/11/88
12/12/88
01/01/89
02/02/89
03/03/89
04/04/89
05/05/89
06/06/89
07/07/89
08/08/89
09/09/89
10/10/89
11/11/89
12/12/89
01/01/90
02/02/90
03/03/90
04/04/90
05/05/90
06/06/90
07/07/90
08/08/90
09/09/90
10/10/90
11/11/90
12/12/90
01/01/91
02/02/91
03/03/91
04/04/91
05/05/91
06/06/91
07/07/91
08/08/91
09/09/91
10/10/91
11/11/91
12/12/91
01/01/92
02/02/92
03/03/92
04/04/92
05/05/92
06/06/92
07/07/92
08/08/92
09/09/92
10/10/92
11/11/92
12/12/92
01/01/93
02/02/93
03/03/93
04/04/93
05/05/93
06/06/93
07/07/93
08/08/93
09/09/93
10/10/93
11/11/93
12/12/93
01/01/94
02/02/94
03/03/94
04/04/94
05/05/94
06/06/94
07/07/94
08/08/94
09/09/94
10/10/94
11/11/94
12/12/94
01/01/95
02/02/95
03/03/95
04/04/95
05/05/95
06/06/95
07/07/95
08/08/95
09/09/95
10/10/95
11/11/95
12/12/95
01/01/96
02/02/96
03/03/96
04/04/96
05/05/96
06/06/96
07/07/96
08/08/96
09/09/96
10/10/96
11/11/96
12/12/9

- ☐ Preliminary Amendment
- ☒ Information Sheet
- ☐ Information Disclosure Statement, PTO-1449 and reference(s)
- ☐ Amend the specification by inserting before the first line the sentence:

--This application claims priority on provisional Application No. _____ filed on _____, the entire contents of which are hereby incorporated by reference.--

- ☒ Other: Power of Attorney regarding Small Entity Statement, ATCC deposit receipts PTA-595, PTA-1161, PTA-1411, CD containing specification.

The filing fee has been calculated as shown below:

			LARGE ENTITY		SMALL ENTITY	
BASIC FEE			\$690.00		\$345.00	
	NUMBER FILED	NUMBER EXTRA	RATE	FEE	RATE	FEE
TOTAL CLAIMS	50- 20 =	30	X 18 =	\$0.00	x 9 =	\$270.00
INDEPENDENT CLAIMS	5- 3 =	2	x 78 =	\$0.00	x 39 =	\$78.00
MULTIPLE DEPENDENT <input type="checkbox"/> CLAIMS PRESENTED			+ \$260.00		+ \$130.00	
TOTAL			\$0.00		\$693.00	

- ☒ The application transmitted herewith is filed in accordance with 37 C.F.R. § 1.41(c). The undersigned has been authorized by the inventor(s) to file the present application. The original duly executed declaration together with the surcharge will be forwarded in due course.
- ☒ A check in the amount of \$693.00 to cover the filing fee is enclosed.

☐ Please charge Deposit Account No. 02-2448 in the amount of \$0.00. A triplicate copy of this transmittal form is enclosed.

☒ Please send correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP or Customer No. 2292
P.O. Box 747
Falls Church, VA 22040-0747
Telephone: (703) 205-8000

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. §§ 1.16 or 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By Raymond C. Stewart #36,623
Raymond C. Stewart, #21,066

RCS/DRN/JDK
2750-942P

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

Attachments

IN THE U.S. PATENT AND TRADEMARK OFFICE

I N F O R M A T I O N S H E E T

Applicant: Nickolai ALEXANDROV, Maxim TROUKHAN
Appl. No.: NEW
Filed: June 16, 2000
For: SEQUENCE-DETERMINED DNA FRAGMENTS AND
 CORRESPONDING POLYPEPTIDES ENCODED
 THEREBY

Priority Claimed: 2750-465P 60/139,763 June 18, 1999

Send Correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP or **CUSTOMER NO. 2292**
P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

The above information is submitted to advise the U.S.P.T.O.
of all relevant facts in connection with the present application.

A timely executed Declaration in accordance with 37 C.F.R.
§ 1.64 will follow.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By *Raymond C. Stewart* #36623
RS Raymond C. Stewart, #21,066

RCS/DRN/JDK
2750-942P

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

STATEMENT CLAIMING SMALL ENTITY STATUS
(37 CFR 1.9(f) & 1.27(c)) - SMALL BUSINESS CONCERN

Docket Number: 2750-942P

Applicant, Patentee, or Identifier: Nickolai ALEXANDROV, Maxim TROUKHAN
Application or Patent No.: NEW Patent Application
Filed or Issued: June 16, 2000
Title: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
ENCODED THEREBY

I hereby state that I am

- ☐ the owner of the small business concern identified below:
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN CERES, INC.

ADDRESS OF SMALL BUSINESS CONCERN 3007 Malibu Canyon Road Malibu, CA 90265

I hereby state that the above identified small business concern qualifies as a small business concern as defined in 37 CFR Part 121 for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby state that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention described in:

- ☒ the specification filed herewith with title as listed above.
☐ the application identified above.
☐ the patent identified above.

If the rights held by the above identified small business concern are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d), or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:

- ☒ no such person, concern, or organization exists.
☐ each such person, concern, or organization is listed below.

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is not longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING Mark J. Nuell (Reg. No. 36, 623)

TITLE IN ORGANIZATION OF PERSON SIGNING Legal Representative of CERES, INC.

ADDRESS OF PERSON SIGNING Birch, Stewart, Kolasch and Birch, LLP.

P.O. Box 747 Falls Church, VA 22040-0747

SIGNATURE

my Nuell

DATE June 16, 2000

**SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING
POLYPEPTIDES ENCODED THEREBY**

This application claims priority under 35 USC §119(c), §119(a-d) and §120 of the following applications, the entire contents of which are hereby incorporated by reference:

Country	Filing Date	Attorney No.	Client No.	Application No.
United States	06/18/99	2750-0465P	00038.001	60/139,763

FIELD OF THE INVENTION

The present invention relates to isolated polynucleotides that represent a complete gene, or a fragment thereof, that is expressed. In addition, the present invention relates to the polypeptide or protein corresponding to the coding sequence of these polynucleotides. The present invention also relates to isolated polynucleotides that represent regulatory regions of genes. The present invention also relates to isolated polynucleotides that represent untranslated regions of genes. The present invention further relates to the use of these isolated polynucleotides and polypeptides and proteins.

DESCRIPTION OF THE RELATED ART

Efforts to map and sequence the genome of a number of organisms are in progress; a few complete genome sequences, for example those of *E. coli* and *Saccharomyces cerevisiae* are known (Blattner et al., *Science* 277:1453 (1997); Goffeau et al., *Science* 274:546 (1996)). The complete genome of a multicellular organism, *C. elegans*, has also been sequenced (See, the *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998)). To date, no complete genome of a plant has been sequenced, nor has a complete cDNA complement of any plant been sequenced.

SUMMARY OF THE INVENTION

The present invention comprises polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, fragments of genes, and/or regulatory elements of genes and/or regions with other functions and/or intergenic regions, hereinafter collectively referred to as Sequence-Determined DNA Fragments (SDFs), from different plant species, particularly corn, wheat, soybean, rice and *Arabidopsis thaliana*, and other plants and or mutants, variants, fragments or fusions of said SDFs and polypeptides or proteins derived therefrom. In some instances, the SDFs span the entirety of a protein-coding segment. In some instances, the entirety of an mRNA is represented. Other objects of the

invention that are also represented by SDFs of the invention are control sequences, such as, but not limited to, promoters. Complements of any sequence of the invention are also considered part of the invention.

Other objects of the invention are polynucleotides comprising exon sequences,
5 polynucleotides comprising intron sequences, polynucleotides comprising introns together with exons, intron/exon junction sequences, 5' untranslated sequences, and 3' untranslated sequences of the SDFs of the present invention. Polynucleotides representing the joinder of any exons described herein, in any arrangement, for example, to produce a sequence encoding any desirable amino acid sequence are within the scope of the invention.

10 The present invention also resides in probes useful for isolating and identifying nucleic acids that hybridize to an SDF of the invention. The probes can be of any length, but more typically are 12-2000 nucleotides in length; more typically, 15 to 200 nucleotides long; even more typically, 18 to 100 nucleotides long.

Yet another object of the invention is a method of isolating and/or identifying nucleic
15 acids using the following steps:

- (a) contacting a probe of the instant invention with a polynucleotide sample under conditions that permit hybridization and formation of a polynucleotide duplex; and
- (b) detecting and/or isolating the duplex of step (a).

20 The conditions for hybridization can be from low to moderate to high stringency conditions. The sample can include a polynucleotide having a sequence unique in a plant genome. Probes and methods of the invention are useful, for example, without limitation, for mapping of genetic traits and/or for positional cloning of a desired fragment of genomic DNA.

Probes and methods of the invention can also be used for detecting alternatively spliced messages within a species. Probes and methods of the invention can further be used to detect or
25 isolate related genes in other plant species using genomic DNA (gDNA) and/or cDNA libraries. In some instances, especially when longer probes and low to moderate stringency hybridization conditions are used; the probe will hybridize to a plurality of cDNA and/or gDNA sequences of a plant. This approach is useful for isolating representatives of gene families which are identifiable by possession of a common functional domain in the gene product or which have
30 common cis-acting regulatory sequences. This approach is also useful for identifying orthologous genes from other organisms.

The present invention also resides in constructs for modulating the expression of the genes comprised of all or a fragment of an SDF. The constructs comprise all or a fragment of

the expressed SDF, or of a complementary sequence. Examples of constructs include ribozymes comprising RNA encoded by an SDF or by a sequence complementary thereto, antisense constructs, constructs comprising coding regions or parts thereof, constructs comprising promoters, introns, untranslated regions, scaffold attachment regions, methylating regions, enhancing or reducing regions, DNA and chromatin conformation modifying sequences, etc. Such constructs can be constructed using viral, plasmid, bacterial artificial chromosomes (BACs), plasmid artificial chromosomes (PACs), autonomous plant plasmids, plant artificial chromosomes or other types of vectors and exist in the plant as autonomous replicating sequences or as DNA integrated into the genome. When inserted into a host cell the construct is, preferably, functionally integrated with, or operatively linked to, a heterologous polynucleotide. For instance, a coding region from an SDF might be operably linked to a promoter that is functional in a plant.

The present invention also resides in host cells, including bacterial or yeast cells or plant cells, and plants that harbor constructs such as described above. Another aspect of the invention relates to methods for modulating expression of specific genes in plants by expression of the coding sequence of the constructs, by regulation of expression of one or more endogenous genes in a plant or by suppression of expression of the polynucleotides of the invention in a plant. Methods of modulation of gene expression include without limitation (1) inserting into a host cell additional copies of a polynucleotide comprising a coding sequence; (2) modulating an endogenous promoter in a host cell; (3) inserting antisense or ribozyme constructs into a host cell and (4) inserting into a host cell a polynucleotide comprising a sequence encoding a variant, fragment, or fusion of the native polypeptides of the instant invention.

BRIEF DESCRIPTION OF THE TABLES

In TABLE 1, the format of the data is as follows:

In Table 1, sequence data are presented in the form of annotation of a reference sequence. The format is shown below. The reference sequence is shown at the top of the annotation file as a 7 digit sequence number preceded by ">" (e.g. >5019261). The sequence identifier is a "gi" number that identifies a specific DNA sequence in the publically accessible BLAST Databases on the NCBI FTP web site (accessible at ncbi.nlm.gov/blast). In particular, the "nt.Z" nucleotide sequence data base at the NCBI FTP site utilizes the "gi" identifiers to assign by NCBI a unique identifier for each sequence in the databases, thereby providing a non-redundant database for sequences from various data bases, including

GenBank, EMBL, DDBJ (DNA Database of Japan) and PDB (Brookhaven Protein Data Bank). Thus, the line in TABLE 1 beginning with sequence number identifies the unique "gi" identifier followed by the corresponding GenBank (gb) accession number and locus. The reference sequence number is followed on the next line by data regarding the length of the sequence ("len") and the number of exons found in the sequence by the analysis program ("nex").

The annotation data are presented in columns; the leftmost column identifies the position of the putative exon in the gene as initial ("init"), internal ("intr") or terminal ("term"). Genes considered composed of a single exon are denoted "sngl". The next column describes the position in the nucleotide sequence beginning the exon ("start") and the next column describes the position in the nucleotide sequence ending the exon ("stop"). The direction of the gene is indicated in the next column, "+" indicating 5' - 3' in the direction presented in the database, "-" indicating the opposite orientation. The "gene number" is given in the final column. Exons having the same gene number are grouped in the order shown to create the relevant coding sequence.

5

>5019261 <= This is the gi number of the public sequence

len = 97208 nex = 121

↑

↑

Length

Number exons

5 of public sequence

	Exon Type	Start	Stop	Direction	Gene Number
10	↓	↓	↓	↓	↓
	Sngl	602	778	+	0
	Sngl	990	1316	+	1
	Sngl	2356	2691	+	2
	Sngl	4634	4735	+	3
15	Sngl	4973	5092	+	4
	Sngl	5746	5874	+	5
	Init	8119	8798	+	6
	Term	9284	9518	+	6
	Init	10827	11150	+	7
20	Term	11294	11335	+	7
	Sngl	12655	12825	+	8
	Sngl	13303	13596	+	9
	Sngl	18654	18782	+	10
	Sngl	19880	20086	+	11
25	Init	21476	21539	+	12
	Intr	21647	21802	+	12
	Term	23488	23567	+	12
	Init	25035	25133	+	13
	Intr	25466	25589	+	13
30	Intr	25677	25786	+	13
	Intr	25899	25962	+	13
	Intr	26045	26109	+	13
	Intr	26188	26253	+	13
	Term	26350	26448	+	13
35	Sngl	27671	27793	+	14
	Sngl	29126	29299	+	15
	Sngl	30266	30364	+	16
	Sngl	31717	31929	+	17
	Sngl	32102	32209	+	18
40	Sngl	32450	32548	+	19

	Sngl	32634	32726	+	20
	Init	35603	35743	+	21
	Term	35829	36185	+	21
	Init	36954	37098	+	22
5	Term	38100	38158	+	22
	Init	39635	39944	+	23
	Intr	40242	40372	+	23
	Intr	40462	40695	+	23
	Intr	40815	41070	+	23
10	Intr	41176	41255	+	23
	Intr	42212	42419	+	23
	Intr	42940	43070	+	23
	Intr	43177	43410	+	23
	Intr	43580	43835	+	23
15	Intr	46672	46715	+	23
	Intr	48334	48532	+	23

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to (I) polynucleotides and methods of use thereof, such as

IA. Probes, Primers and Substrates;

IB. Methods of Detection and Isolation;

B.1. Hybridization;

B.2. Methods of Mapping;

B.3. Southern Blotting;

B.4. Isolating cDNA from Related Organisms;

B.5. Isolating and/or Identifying Orthologous Genes

IC. Methods of Inhibiting Gene Expression

C.1. Antisense

C.2. Ribozyme Constructs;

C.3. Chimeraplasts;

C.4. Co-Suppression;

C.5. Transcriptional Silencing

C.6. Other Methods to Inhibit Gene Expression

ID. Methods of Functional Analysis;

IE. Promoter Sequences and Their Use;

IF. UTRs and/or Intron Sequences and Their Use; and

IG. Coding Sequences and Their Use.

The invention also relates to (II) polypeptides and proteins and methods of use thereof, such as IIA. Native Polypeptides and Proteins

A.1 Antibodies

A.2 In Vitro Applications

IIB. Polypeptide Variants, Fragments and Fusions

B.1 Variants

B.2 Fragments

B.3 Fusions

The invention also includes (III) methods of modulating polypeptide production, such as

IIIA. Suppression

A.1 Antisense

A.2 Ribozymes

A.3 Co-suppression

A.4 Insertion of Sequences into the Gene to be Modulated

A.5 Promoter Modulation

A.6 Expression of Genes containing Dominant-Negative Mutations

IIIB. Enhanced Expression

B.1 Insertion of an Exogenous Gene

B.2 Promoter Modulation

The invention further concerns (IV) gene constructs and vector construction, such as

IVA. Coding Sequences

IVB. Promoters

IVC. Signal Peptides

The invention still further relates to

V Transformation Techniques

Definitions

Allelic variant An "allelic variant" is an alternative form of the same SDF, which resides at the same chromosomal locus in the organism. Allelic variations can occur in any portion of the gene sequence, including regulatory regions. Allelic variants can arise by normal genetic variation in a population. Allelic variants can also be produced by genetic engineering methods. An allelic variant can be one that is found in a naturally occurring plant, including a cultivar or ecotype. An allelic variant may or may not give rise to a phenotypic change, and may or may not be expressed. An allele can result in a detectable change in the phenotype of the trait represented by the locus. A phenotypically silent allele can give rise to a product.

Alternatively spliced messages Within the context of the current invention, "alternatively spliced messages" refers to mature mRNAs originating from a single gene with variations in the number and/or identity of exons, introns and/or intron-exon junctions.

Chimeric The term "chimeric" is used to describe genes, as defined supra, or constructs wherein at least two of the elements of the gene or construct, such as the promoter and the coding sequence and/or other regulatory sequences and/or filler sequences and/or complements thereof, are heterologous to each other.

Constitutive Promoter: Promoters referred to herein as "constitutive promoters" actively promote transcription under most, but not necessarily all, environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcript initiation region and the 1' or 2' promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes, such as the maize ubiquitin-1 promoter, known to those of skill.

Coordinately Expressed: The term "coordinately expressed," as used in the current invention, refers to genes that are expressed at the same or a similar time and/or stage and/or under the same or similar environmental conditions.

Domain: Domains are fingerprints or signatures that can be used to characterize protein families and/or parts of proteins. Such fingerprints or signatures can comprise conserved (1) primary sequence, (2) secondary structure, and/or (3) three-dimensional conformation. Generally, each domain has been associated with either a family of proteins or motifs. Typically, these families and/or motifs have been correlated with specific *in-vitro* and/or *in-vivo* activities. A domain can be any length, including the entirety of the sequence of a protein. Detailed descriptions of the domains, associated families and motifs, and correlated activities of the polypeptides of the instant invention are described below. Usually, the polypeptides with designated domain(s) can exhibit at least one activity that is exhibited by any polypeptide that comprises the same domain(s).

Endogenous The term “endogenous,” within the context of the current invention refers to any polynucleotide, polypeptide or protein sequence which is a natural part of a cell or organisms regenerated from said cell.

Exogenous “Exogenous,” as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation (of dicots - e.g. Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983); of monocots, representative papers are those by Escudero et al., *Plant J.* 10:355 (1996), Ishida et al., *Nature Biotechnology* 14:745 (1996), May et al., *Bio/Technology* 13:486 (1995)), biolistic methods (Armaleo et al., *Current Genetics* 17:97 1990)), electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T₀ for the primary transgenic plant and T₁ for the first generation. The term “exogenous” as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

Filler sequence: As used herein, “filler sequence” refers to any nucleotide sequence that is inserted into DNA construct to evoke a particular spacing between particular components

such as a promoter and a coding region and may provide an additional attribute such as a restriction enzyme site.

Gene: The term “gene,” as used in the context of the current invention, encompasses all regulatory and coding sequence contiguously associated with a single hereditary unit with a genetic function (see SCHEMATIC 1). Genes can include non-coding sequences that modulate the genetic function that include, but are not limited to, those that specify polyadenylation, transcriptional regulation, DNA conformation, chromatin conformation, extent and position of base methylation and binding sites of proteins that control all of these.

Genes comprised of “exons” (coding sequences), which may be interrupted by “introns” (non-coding sequences), encode proteins. A gene’s genetic function may require only RNA expression or protein production, or may only require binding of proteins and/or nucleic acids without associated expression. In certain cases, genes adjacent to one another may share sequence in such a way that one gene will overlap the other. A gene can be found within the genome of an organism, artificial chromosome, plasmid, vector, etc., or as a separate isolated entity.

Gene Family: “Gene family” is used in the current invention to describe a group of functionally related genes, each of which encodes a separate protein.

Heterologous sequences: “Heterologous sequences” are those that are not operatively linked or are not contiguous to each other in nature. For example, a promoter from corn is considered heterologous to an *Arabidopsis* coding region sequence. Also, a promoter from a gene encoding a growth factor from corn is considered heterologous to a sequence encoding the corn receptor for the growth factor. Regulatory element sequences, such as UTRs or 3’ end termination sequences that do not originate in nature from the same gene as the coding sequence originates from, are considered heterologous to said coding sequence. Elements operatively linked in nature and contiguous to each other are not heterologous to each other. On the other hand, these same elements remain operatively linked but become heterologous if other filler sequence is placed between them. Thus, the promoter and coding sequences of a corn gene expressing an amino acid transporter are not heterologous to each other, but the promoter and coding sequence of a corn gene operatively linked in a novel manner are heterologous.

Homologous gene In the current invention, "homologous gene" refers to a gene that shares sequence similarity with the gene of interest. This similarity may be in only a fragment of the sequence and often represents a functional domain such as, examples including without limitation a DNA binding domain, a domain with tyrosine kinase activity, or the like. The functional activities of homologous genes are not necessarily the same.

Inducible Promoter An "inducible promoter" in the context of the current invention refers to a promoter which is regulated under certain conditions, such as light, chemical concentration, protein concentration, conditions in an organism, cell, or organelle, etc. A typical example of an inducible promoter, which can be utilized with the polynucleotides of the present invention, is PARSK1, the promoter from the *Arabidopsis* gene encoding a serine-threonine kinase enzyme, and which promoter is induced by dehydration, abscissic acid and sodium chloride (Wang and Goodman, *Plant J.* 8:37 (1995)) Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

Intergenic region "Intergenic region," as used in the current invention, refers to nucleotide sequence occurring in the genome that separates adjacent genes.

Mutant gene In the current invention, "mutant" refers to a heritable change in DNA sequence at a specific location. Mutants of the current invention may or may not have an associated identifiable function when the mutant gene is transcribed.

Orthologous Gene In the current invention "orthologous gene" refers to a second gene that encodes a gene product that performs a similar function as the product of a first gene. The orthologous gene may also have a degree of sequence similarity to the first gene. The orthologous gene may encode a polypeptide that exhibits a degree of sequence similarity to a polypeptide corresponding to a first gene. The sequence similarity can be found within a functional domain or along the entire length of the coding sequence of the genes and/or their corresponding polypeptides.

Percentage of sequence identity "Percentage of sequence identity," as used herein, is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of the polynucleotide or amino acid sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection. Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. The term "substantial sequence identity" between polynucleotide or polypeptide sequences refers to polynucleotide or polypeptide comprising a sequence that has at least 80% sequence identity, preferably at least 85%, more preferably at least 90% and most preferably at least 95%, even more preferably, at least 96%, 97%, 98% or 99% sequence identity compared to a reference sequence using the programs.

Plant Promoter A "plant promoter" is a promoter capable of initiating transcription in plant cells and can drive or facilitate transcription of a fragment of the SDF of the instant invention or a coding sequence of the SDF of the instant invention. Such promoters need not be of plant origin. For example, promoters derived from plant viruses, such as the CaMV35S promoter or from *Agrobacterium tumefaciens* such as the T-DNA promoters, can be plant promoters. A typical example of a plant promoter of plant origin is the maize ubiquitin-1 (ubi-1) promoter known to those of skill.

Promoter: The term "promoter," as used herein, refers to a region of sequence determinants located upstream from the start of transcription of a gene and which are involved in recognition and binding of RNA polymerase and other proteins to initiate and modulate transcription. A basal promoter is the minimal sequence necessary for assembly of a transcription complex required for transcription initiation. Basal promoters frequently include a "TATA box" element usually located between 15 and 35 nucleotides upstream from the site of initiation of transcription. Basal promoters also sometimes include a "CCAAT box" element (typically a sequence CCAAT) and/or a GGCG sequence, usually located between 40 and 200 nucleotides, preferably 60 to 120 nucleotides, upstream from the start site of transcription.

Public sequence: The term "public sequence," as used in the context of the instant application, refers to any sequence that has been deposited in a publicly accessible database. This term encompasses both amino acid and nucleotide sequences. Such sequences are publicly accessible, for example, on the BLAST databases on the NCBI FTP web site (accessible at ncbi.nlm.gov/blast). The database at the NCBI GTP site utilizes "gi" numbers assigned by NCBI as a unique identifier for each sequence in the databases, thereby providing a non-redundant database for sequence from various databases, including GenBank, EMBL, DDBJ, (DNA Database of Japan) and PDB (Brookhaven Protein Data Bank).

Regulatory Sequence The term "regulatory sequence," as used in the current invention, refers to any nucleotide sequence that influences transcription or translation initiation and rate, and stability and/or mobility of the transcript or polypeptide product.

Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start site, termination sequence, polyadenylation sequence, introns, certain sequences within a coding sequence, etc.

Related Sequences: "Related sequences" refer to either a polypeptide or a nucleotide sequence that exhibits some degree of sequence similarity with a sequence described in Table 1.

Scaffold Attachment Region (SAR) As used herein, "scaffold attachment region" is a DNA sequence that anchors chromatin to the nuclear matrix or scaffold to generate loop domains that can have either a transcriptionally active or inactive structure (Spiker and Thompson (1996) *Plant Physiol.* 110: 15-21).

Sequence-determined DNA fragments (SDFs) "Sequence-determined DNA fragments" as used in the current invention are isolated sequences of genes, fragments of genes, intergenic regions or contiguous DNA from plant genomic DNA or cDNA or RNA the sequence of which has been determined.

Signal Peptide A "signal peptide" as used in the current invention is an amino acid sequence that targets the protein for secretion, for transport to an intracellular compartment or organelle or for incorporation into a membrane. Signal peptides are indicated in the tables and a more detailed description located below.

Specific Promoter In the context of the current invention, "specific promoters" refers to a subset of inducible promoters that have a high preference for being induced in a specific tissue or cell and/or at a specific time during development of an organism. By "high preference" is meant at least 3-fold, preferably 5-fold, more preferably at least 10-fold still more preferably at least 20-fold, 50-fold or 100-fold increase in transcription in the desired tissue over the transcription in any other tissue. Typical examples of temporal and/or tissue specific promoters of plant origin that can be used with the polynucleotides of the present invention, are: PTA29, a promoter which is capable of driving gene transcription specifically in tapetum and only during anther development (Koltonow et al., *Plant Cell* 2:1201 (1990); RCc2 and RCc3, promoters that direct root-specific gene transcription in rice (Xu et al., *Plant Mol. Biol.* 27:237 (1995); TobRB27, a root-specific promoter from tobacco (Yamamoto et al., *Plant Cell* 3:371 (1991)). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues or organs, such as root, ovule, fruit, seeds, or flowers. Other suitable promoters include those from genes encoding storage proteins or the lipid body membrane protein, oleosin. A few root-specific promoters are noted above.

Stringency "Stringency" as used herein is a function of probe length, probe composition (G + C content), and salt concentration, organic solvent concentration, and temperature of

hybridization or wash conditions. Stringency is typically compared by the parameter T_m , which is the temperature at which 50% of the complementary molecules in the hybridization are hybridized, in terms of a temperature differential from T_m . High stringency conditions are those providing a condition of $T_m - 5^\circ\text{C}$ to $T_m - 10^\circ\text{C}$. Medium or moderate stringency conditions are those providing $T_m - 20^\circ\text{C}$ to $T_m - 29^\circ\text{C}$. Low stringency conditions are those providing a condition of $T_m - 40^\circ\text{C}$ to $T_m - 48^\circ\text{C}$. The relationship of hybridization conditions to T_m (in $^\circ\text{C}$) is expressed in the mathematical equation

$$T_m = 81.5 - 16.6(\log_{10}[\text{Na}^+]) + 0.41(\%G+C) - (600/N) \quad (1)$$

where N is the length of the probe. This equation works well for probes 14 to 70 nucleotides in length that are identical to the target sequence. The equation below for T_m of DNA-DNA hybrids is useful for probes in the range of 50 to greater than 500 nucleotides, and for conditions that include an organic solvent (formamide).

$$T_m = 81.5 + 16.6 \log \{ [\text{Na}^+]/(1 + 0.7[\text{Na}^+]) \} + 0.41(\%G+C) - 500/L - 0.63(\%\text{formamide}) \quad (2)$$

where L is the length of the probe in the hybrid. (P. Tijessen, "Hybridization with Nucleic Acid Probes" in Laboratory Techniques in Biochemistry and Molecular Biology, P.C. van der Vliet, ed., c. 1993 by Elsevier, Amsterdam.) The T_m of equation (2) is affected by the nature of the hybrid; for DNA-RNA hybrids T_m is $10-15^\circ\text{C}$ higher than calculated, for RNA-RNA hybrids T_m is $20-25^\circ\text{C}$ higher. Because the T_m decreases about 1°C for each 1% decrease in homology when a long probe is used (Bonner et al., *J. Mol. Biol.* 81:123 (1973)), stringency conditions can be adjusted to favor detection of identical genes or related family members.

Equation (2) is derived assuming equilibrium and therefore, hybridizations according to the present invention are most preferably performed under conditions of probe excess and for sufficient time to achieve equilibrium. The time required to reach equilibrium can be shortened by inclusion of a hybridization accelerator such as dextran sulfate or another high volume polymer in the hybridization buffer.

Stringency can be controlled during the hybridization reaction or after hybridization has occurred by altering the salt and temperature conditions of the wash solutions used. The formulas shown above are equally valid when used to compute the stringency of a wash

solution. Preferred wash solution stringencies lie within the ranges stated above; high stringency is 5-8°C below T_m , medium or moderate stringency is 26-29°C below T_m and low stringency is 45-48°C below T_m .

- 5 Substantially free of A A composition containing A is “substantially free of “ B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight. For example, a plant gene or DNA sequence can be considered substantially free of other plant genes or DNA sequences.

10

Translational start site In the context of the current invention, a “translational start site” is usually an ATG in the cDNA transcript, more usually the first ATG. A single cDNA, however, may have multiple translational start sites.

15

Transcription start site “Transcription start site” is used in the current invention to describe the point at which transcription is initiated. This point is typically located about 25 nucleotides downstream from a TFIID binding site, such as a TATA box. Transcription can initiate at one or more sites within the gene, and a single gene may have multiple transcriptional start sites, some of which may be specific for transcription in a particular cell-type or tissue.

20

Untranslated region (UTR) A “UTR” is any contiguous series of nucleotide bases that is transcribed, but is not translated. These untranslated regions may be associated with particular functions such as increasing mRNA message stability. Examples of UTRs include, but are not limited to polyadenylation signals, terminations sequences, sequences located between the transcriptional start site and the first exon (5' UTR) and sequences located between the last exon and the end of the mRNA (3' UTR).

25

Variant: The term “variant” is used herein to denote a polypeptide or protein or polynucleotide molecule that differs from others of its kind in some way. For example, polypeptide and protein variants can consist of changes in amino acid sequence and/or charge and/or post-translational modifications (such as glycosylation, etc).

30

055326-061600

DETAILED DESCRIPTION OF THE INVENTION**I. Polynucleotides**

Exemplified SDFs of the invention represent fragments of the genome of corn, wheat, rice, soybean or *Arabidopsis* and/or represent mRNA expressed from that genome. The isolated nucleic acid of the invention also encompasses corresponding fragments of the genome and/or cDNA complement of other organisms as described in detail below.

Polynucleotides of the invention can be isolated from polynucleotide libraries using primers comprising sequence similar to those described by Table 1. See, for example, the methods described in Sambrook et al., supra.

Alternatively, the polynucleotides of the invention can be produced by chemical synthesis. Such synthesis methods are described below.

It is contemplated that the nucleotide sequences presented herein may contain some small percentage of errors. These errors may arise in the normal course of determination of nucleotide sequences. Sequence errors can be corrected by obtaining seeds deposited under the accession numbers cited herein, propagating them, isolating genomic DNA or appropriate mRNA from the resulting plants or seeds thereof, amplifying the relevant fragment of the genomic DNA or mRNA using primers having a sequence that flanks the erroneous sequence, and sequencing the amplification product.

I.A. Probes, Primers and Substrates

SDFs of the invention can be applied to substrates for use in array applications such as, but not limited to, assays of global gene expression, for example under varying conditions of development, growth conditions. The arrays can also be used in diagnostic or forensic methods (WO95/35505, US 5,445,943 and US 5,410,270).

Probes and primers of the instant invention will hybridize to a polynucleotide comprising a sequence in Table 1. Though many different nucleotide sequences can encode an amino acid sequence, the sequences of Table 1 are generally preferred for encoding polypeptides of the invention. However, the sequence of the probes and/or primers of the instant invention need not be identical to those in Table 1 or the complements thereof. For example, some variation in probe or primer sequence and/or length can allow additional

family members to be detected, as well as orthologous genes and more taxonomically distant related sequences. Similarly, probes and/or primers of the invention can include additional nucleotides that serve as a label for detecting the formed duplex or for subsequent cloning purposes.

Probe length will vary depending on the application. For use as primers, probes are 12-40 nucleotides, preferably 18-30 nucleotides long. For use in mapping, probes are preferably 50 to 500 nucleotides, preferably 100-250 nucleotides long. For Southern hybridizations, probes as long as several kilobases can be used as explained below.

The probes and/or primers can be produced by synthetic procedures such as the triester method of Matteucci et al. *J. Am. Chem. Soc.* **103**:3185(1981); or according to Urdea et al. *Proc. Natl. Acad.* **80**:7461 (1981) or using commercially available automated oligonucleotide synthesizers.

I.B. Methods of Detection and Isolation

The polynucleotides of the invention can be utilized in a number of methods known to those skilled in the art as probes and/or primers to isolate and detect polynucleotides, including, without limitation: Southern, Northern, Branched DNA hybridization assays, polymerase chain reaction, and microarray assays, and variations thereof. Specific methods given by way of examples, and discussed below include:

Hybridization

Methods of Mapping

Southern Blotting

Isolating cDNA from Related Organisms

Isolating and/or Identifying Orthologous Genes.

Also, the nucleic acid molecules of the invention can be used in other methods, such as high density oligonucleotide hybridizing assays, described, for example, in U.S. Pat. Nos. 6,004,753; 5,945,306; 5,945,287; 5,945,308; 5,919,686; 5,919,661; 5,919,627; 5,874,248; 5,871,973; 5,871,971; and 5,871,930; and PCT Pub. Nos. WO 9946380; WO 9933981; WO 9933870; WO 9931252; WO 9915658; WO 9906572; WO 9858052; WO 9958672; and WO 9810858.

B.1. Hybridization

The isolated SDFs of Table 1 of the present invention can be used as probes and/or primers for detection and/or isolation of related polynucleotide sequences through hybridization. Hybridization of one nucleic acid to another constitutes a physical property that defines the subject SDF of the invention and the identified related sequences. Also, such hybridization imposes structural limitations on the pair. A good general discussion of the factors for determining hybridization conditions is provided by Sambrook et al. ("Molecular Cloning, a Laboratory Manual, 2nd ed., c. 1989 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; *see esp.*, chapters 11 and 12). Additional considerations and details of the physical chemistry of hybridization are provided by G.H. Keller and M.M. Manak "DNA Probes", 2nd Ed. pp. 1-25, c. 1993 by Stockton Press, New York, NY.

Depending on the stringency of the conditions under which these probes and/or primers are used, polynucleotides exhibiting a wide range of similarity to those in Table 1 can be detected or isolated. When the practitioner wishes to examine the result of membrane hybridizations under a variety of stringencies, an efficient way to do so is to perform the hybridization under a low stringency condition, then to wash the hybridization membrane under increasingly stringent conditions.

When using SDFs to identify orthologous genes in other species, the practitioner will preferably adjust the amount of target DNA of each species so that, as nearly as is practical, the same number of genome equivalents are present for each species examined. This prevents faint signals from species having large genomes, and thus small numbers of genome equivalents per mass of DNA, from erroneously being interpreted as absence of the corresponding gene in the genome.

The probes and/or primers of the instant invention can also be used to detect or isolate nucleotides that are "identical" to the probes or primers. Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below.

Isolated polynucleotides within the scope of the invention also include allelic variants of the specific sequences presented in Table 1. The probes and/or primers of the invention can also be used to detect and/or isolate polynucleotides exhibiting at least 80% sequence identity with the sequences of Table 1 or fragments thereof.

With respect to nucleotide sequences, degeneracy of the genetic code provides the possibility to substitute at least one base of the base sequence of a gene with a different base without causing the amino acid sequence of the polypeptide produced from the gene to be changed. Hence, the DNA of the present invention may also have any base sequence that has been changed from a sequence in Table 1 by substitution in accordance with degeneracy of genetic code. References describing codon usage include: Carels *et al.*, *J. Mol. Evol.* **46**: 45 (1998) and Fennoy *et al.*, *Nucl. Acids Res.* **21(23)**: 5294 (1993).

B.2. Mapping

The isolated SDF DNA of the invention can be used to create various types of genetic and physical maps of the genome of corn, Arabidopsis, soybean, rice, wheat, or other plants. Some SDFs may be absolutely associated with particular phenotypic traits, allowing construction of gross genetic maps. While not all SDFs will immediately be associated with a phenotype, all SDFs can be used as probes for identifying polymorphisms associated with phenotypes of interest. Briefly, one method of mapping involves total DNA isolation from individuals. It is subsequently cleaved with one or more restriction enzymes, separated according to mass, transferred to a solid support, hybridized with SDF DNA and the pattern of fragments compared. Polymorphisms associated with a particular SDF are visualized as differences in the size of fragments produced between individual DNA samples after digestion with a particular restriction enzyme and hybridization with the SDF. After identification of polymorphic SDF sequences, linkage studies can be conducted. By using the individuals showing polymorphisms as parents in crossing programs, F2 progeny recombinants or recombinant inbreds, for example, are then analyzed. The order of DNA polymorphisms along the chromosomes can be determined based on the frequency with which they are inherited together versus independently. The closer two polymorphisms are together in a chromosome the higher the probability that they are inherited together. Integration of the relative positions of all the polymorphisms and associated marker SDFs can produce a genetic map of the species, where the distances between markers reflect the recombination frequencies in that chromosome segment.

The use of recombinant inbred lines for such genetic mapping is described for *Arabidopsis* by Alonso-Blanco *et al.* (*Methods in Molecular Biology*, vol.82, "Arabidopsis Protocols", pp. 137-146, J.M. Martinez-Zapater and J. Salinas, eds., c. 1998 by Humana Press, Totowa, NJ) and for corn by Burr ("Mapping Genes with Recombinant Inbreds", pp.

249-254. In Freeling, M. and V. Walbot (Ed.), *The Maize Handbook*, c. 1994 by Springer-Verlag New York, Inc.: New York, NY, USA; Berlin Germany; Burr et al. *Genetics* (1998) 118: 519; Gardiner, J. et al., (1993) *Genetics* 134: 917). This procedure, however, is not limited to plants and can be used for other organisms (such as yeast) or for individual cells.

5 The SDFs of the present invention can also be used for simple sequence repeat (SSR) mapping. Rice SSR mapping is described by Morgante et al. (*The Plant Journal* (1993) 3: 165), Panaud et al. (*Genome* (1995) 38: 1170); Senior et al. (*Crop Science* (1996) 36: 1676), Taramino et al. (*Genome* (1996) 39: 277) and Ahn et al. (*Molecular and General Genetics* (1993) 241: 483-90). SSR mapping can be achieved using various methods. In one instance, 10 polymorphisms are identified when sequence specific probes contained within an SDF flanking an SSR are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals of interest. Here, a change in the number of tandem repeats between the SSR-flanking sequences produces differently sized fragments (U.S. Patent 5,766,847). Alternatively, polymorphisms can be identified by using the PCR 15 fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (U.H. Refseth et al., (1997) *Electrophoresis* 18: 1519).

Genetic and physical maps of crop species have many uses. For example, these maps can be used to devise positional cloning strategies for isolating novel genes from the mapped crop species. In addition, because the genomes of closely related species are largely syntenic 20 (that is, they display the same ordering of genes within the genome), these maps can be used to isolate novel alleles from relatives of crop species by positional cloning strategies.

The various types of maps discussed above can be used with the SDFs of the invention to identify Quantitative Trait Loci (QTLs). Many important crop traits, such as the solids content of tomatoes, are quantitative traits and result from the combined interactions of 25 several genes. These genes reside at different loci in the genome, oftentimes on different chromosomes, and generally exhibit multiple alleles at each locus. The SDFs of the invention can be used to identify QTLs and isolate specific alleles as described by de Vicente and Tanksley (*Genetics* 134:585 (1993)). In addition to isolating QTL alleles in present crop species, the SDFs of the invention can also be used to isolate alleles from the corresponding 30 QTL of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once a desired allele combination has been identified, crop improvement can be accomplished either through

biotechnological means or by directed conventional breeding programs (for review see Tanksley and McCouch, *Science* 277:1063 (1997)).

In another embodiment, the SDFs can be used to help create physical maps of the genome of corn, *Arabidopsis* and related species. Where SDFs have been ordered on a genetic map, as described above, they can be used as probes to discover which clones in large libraries of plant DNA fragments in YACs, BACs, etc. contain the same SDF or similar sequences, thereby facilitating the assignment of the large DNA fragments to chromosomal positions. Subsequently, the large BACs, YACs, etc. can be ordered unambiguously by more detailed studies of their sequence composition (e.g. Marra et al. (1997) Genomic Research 7:1072-1084) and by using their end or other sequences to find the identical sequences in other cloned DNA fragments. The overlapping of DNA sequences in this way allows large contigs of plant sequences to be built that, when sufficiently extended, provide a complete physical map of a chromosome. Sometimes the SDFs themselves will provide the means of joining cloned sequences into a contig.

The patent publication WO95/35505 and U.S. Patents 5,445,943 and 5,410,270 describe scanning multiple alleles of a plurality of loci using hybridization to arrays of oligonucleotides. These techniques are useful for each of the types of mapping discussed above.

Following the procedures described above and using a plurality of the SDFs of the present invention, any individual can be genotyped. These individual genotypes can be used for the identification of particular cultivars, varieties, lines, ecotypes and genetically modified plants or can serve as tools for subsequent genetic studies involving multiple phenotypic traits.

B.3 Southern Blot Hybridization

The sequences from Table 1 can be used as probes for various hybridization techniques. These techniques are useful for detecting target polynucleotides in a sample or for determining whether transgenic plants, seeds or host cells harbor a gene or sequence of interest and thus might be expected to exhibit a particular trait or phenotype.

In addition, the SDFs from the invention can be used to isolate additional members of gene families from the same or different species and/or orthologous genes from the same or different species. This is accomplished by hybridizing an SDF to, for example, a Southern blot containing the appropriate genomic DNA or cDNA. Given the resulting hybridization

data, one of ordinary skill in the art could distinguish and isolate the correct DNA fragments by size, restriction sites, sequence and stated hybridization conditions from a gel or from a library.

Identification and isolation of orthologous genes from closely related species and alleles within a species is particularly desirable because of their potential for crop improvement. Many important crop traits, such as the solid content of tomatoes, result from the combined interactions of the products of several genes residing at different loci in the genome. Generally, alleles at each of these loci can make quantitative differences to the trait. By identifying and isolating numerous alleles for each locus from within or different species, transgenic plants with various combinations of alleles can be created and the effects of the combinations measured. Once a more favorable allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs (Tanksley et al. *Science* 277:1063(1997)).

The results from hybridizations of the SDFs of the invention to, for example, Southern blots containing DNA from another species can also be used to generate restriction fragment maps for the corresponding genomic regions. These maps provide additional information about the relative positions of restriction sites within fragments, further distinguishing mapped DNA from the remainder of the genome.

Physical maps can be made by digesting genomic DNA with different combinations of restriction enzymes.

Probes for Southern blotting to distinguish individual restriction fragments can range in size from 15 to 20 nucleotides to several thousand nucleotides. More preferably, the probe is 100 to 1,000 nucleotides long for identifying members of a gene family when it is found that repetitive sequences would complicate the hybridization. For identifying an entire corresponding gene in another species, the probe is more preferably the length of the gene, typically 2,000 to 10,000 nucleotides, but probes 50-1,000 nucleotides long might be used. Some genes, however, might require probes up to 1,500 nucleotides long or overlapping probes constituting the full-length sequence to span their lengths.

Also, while it is preferred that the probe be homogeneous with respect to its sequence, it is not necessary. For example, as described below, a probe representing members of a gene family having diverse sequences can be generated using PCR to amplify genomic DNA or

RNA templates using primers derived from SDFs that include sequences that define the gene family.

For identifying corresponding genes in another species, the next most preferable probe is a cDNA spanning the entire coding sequence, which allows all of the mRNA-coding fragment of the gene to be identified. Probes for Southern blotting can easily be generated from SDFs by making primers having the sequence at the ends of the SDF and using corn or *Arabidopsis* genomic DNA as a template. In instances where the SDF includes sequence conserved among species, primers including the conserved sequence can be used for PCR with genomic DNA from a species of interest to obtain a probe.

Similarly, if the SDF includes a domain of interest, that fragment of the SDF can be used to make primers and, with appropriate template DNA, used to make a probe to identify genes containing the domain. Alternatively, the PCR products can be resolved, for example by gel electrophoresis, and cloned and/or sequenced. Using Southern hybridization, the variants of the domain among members of a gene family, both within and across species, can be examined.

B.4.1 Isolating DNA from Related Organisms

The SDFs of the invention can be used to isolate the corresponding DNA from other organisms. Either cDNA or genomic DNA can be isolated. For isolating genomic DNA, a lambda, cosmid, BAC or YAC, or other large insert genomic library from the plant of interest can be constructed using standard molecular biology techniques as described in detail by Sambrook et al. 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York) and by Ausubel et al. 1992 (Current Protocols in Molecular Biology, Greene Publishing, New York).

To screen a phage library, for example, recombinant lambda clones are plated out on appropriate bacterial medium using an appropriate *E. coli* host strain. The resulting plaques are lifted from the plates using nylon or nitrocellulose filters. The plaque lifts are processed through denaturation, neutralization, and washing treatments following the standard protocols outlined by Ausubel et al. (1992). The plaque lifts are hybridized to either radioactively labeled or non-radioactively labeled SDF DNA at room temperature for about 16 hours, usually in the presence of 50% formamide and 5X SSC (sodium chloride and sodium citrate) buffer and blocking reagents. The plaque lifts are then washed at 42°C with 1% Sodium Dodecyl Sulfate (SDS) and at a particular concentration of SSC. The SSC concentration used

is dependent upon the stringency at which hybridization occurred in the initial Southern blot analysis performed. For example, if a fragment hybridized under medium stringency (e.g., $T_m - 20^\circ\text{C}$), then this condition is maintained or preferably adjusted to a less stringent condition (e.g., $T_m - 30^\circ\text{C}$) to wash the plaque lifts. Positive clones show detectable hybridization e.g., by exposure to X-ray films or chromogen formation. The positive clones are then subsequently isolated for purification using the same general protocol outlined above. Once the clone is purified, restriction analysis can be conducted to narrow the region corresponding to the gene of interest. The restriction analysis and succeeding subcloning steps can be done using procedures described by, for example Sambrook et al. (1989) cited above.

The procedures outlined for the lambda library are essentially similar to those used for YAC library screening, except that the YAC clones are harbored in bacterial colonies. The YAC clones are plated out at reasonable density on nitrocellulose or nylon filters supported by appropriate bacterial medium in petri plates. Following the growth of the bacterial clones, the filters are processed through the denaturation, neutralization, and washing steps following the procedures of Ausubel et al. 1992. The same hybridization procedures for lambda library screening are followed.

To isolate cDNA, similar procedures using appropriately modified vectors are employed. For instance, the library can be constructed in a lambda vector appropriate for cloning cDNA such as $\lambda\text{gt}11$. Alternatively, the cDNA library can be made in a plasmid vector. cDNA for cloning can be prepared by any of the methods known in the art, but is preferably prepared as described above. Preferably, a cDNA library will include a high proportion of full-length clones.

B. 5. Isolating and/or Identifying Orthologous Genes

Probes and primers of the invention can be used to identify and/or isolate polynucleotides related to those in Table 1. Related polynucleotides are those that are native to other plant organisms and exhibit either similar sequence or encode polypeptides with similar biological activity. One specific example is an orthologous gene. Orthologous genes have the same functional activity. As such, orthologous genes may be distinguished from homologous genes. The percentage of identity is a function of evolutionary separation and, in closely related species, the percentage of identity can be 98 to 100%. The amino acid sequence of a protein encoded by an orthologous gene can be less than 75% identical, but tends to be at least 75% or at

least 80% identical, more preferably at least 90%, most preferably at least 95% identical to the amino acid sequence of the reference protein.

To find orthologous genes, the probes are hybridized to nucleic acids from a species of interest under low stringency conditions, preferably one where sequences containing as much as 40-45% mismatches will be able to hybridize. This condition is established by $T_m - 40^\circ\text{C}$ to $T_m - 48^\circ\text{C}$ (see below). Blots are then washed under conditions of increasing stringency. It is preferable that the wash stringency be such that sequences that are 85 to 100% identical will hybridize. More preferably, sequences 90 to 100% identical will hybridize and most preferably only sequences greater than 95% identical will hybridize. One of ordinary skill in the art will recognize that, due to degeneracy in the genetic code, amino acid sequences that are identical can be encoded by DNA sequences as little as 67% identical or less. Thus, it is preferable, for example, to make an overlapping series of shorter probes, on the order of 24 to 45 nucleotides, and individually hybridize them to the same arrayed library to avoid the problem of degeneracy introducing large numbers of mismatches.

As evolutionary divergence increases, genome sequences also tend to diverge. Thus, one of skill will recognize that searches for orthologous genes between more divergent species will require the use of lower stringency conditions compared to searches between closely related species. Also, degeneracy of the genetic code is more of a problem for searches in the genome of a species more distant evolutionarily from the species that is the source of the SDF probe sequences.

Therefore the method described in Bouckaert et al., U.S. Ser. No. 60/121,700 Atty. Dkt. No. 2750-117P, Client Dkt. No. 00010.001, filed February 25, 1999, hereby incorporated in its entirety by reference, can be applied to the SDFs of the present invention to isolate related genes from plant species which do not hybridize to the corn *Arabidopsis*, soybean, rice, wheat, and other plant sequences of Table 1.

Identification of the relationship of nucleotide or amino acid sequences among plant species can be done by comparing the nucleotide or amino acid sequences of SDFs of the present application with nucleotide or amino acid sequences of other SDFs such as those present in applications listed in the table below:

Attorney Docket	Client Docket	Filing Date	Application
2750-0301P	80002.001	9/4/1998	60/099,672
2750-0300P	80001.001	9/4/1998	60/099,671
2750-0302P	80003.001	9/11/1998	60/099,933
2750-0304P	80004.001	9/17/1998	60/100,864
2750-0305P	80005.001	9/18/1998	60/101,042

Attorney Docket	Client Docket	Filing Date	Application
2750-0306P	80006.001	9/21/1998	60/101,255
2750-0307P	80007.001	9/24/1998	60/101,682
2750-0308P	80008.001	9/30/1998	60/102,533
2750-0309P	80009.001	9/30/1998	60/102,460
2750-0310P	80010.001	10/5/1998	60/103,116
2750-0311P	80011.001	10/5/1998	60/103,141
2750-0312P	80012.001	10/6/1998	60/103,215
2750-0313P	80013.001	10/8/1998	60/103,554
2750-0314P	80014.001	10/9/1998	60/103,574
2750-0315P	80015.001	10/13/1998	60/103,907
2750-0316P	80016.001	10/14/1998	60/104,268
2750-0317P	80017.001	10/16/1998	60/104,680
2750-0318P	80018.001	10/19/1998	60/104,828
2750-0319P	80019.001	10/20/1998	60/105,008
2750-0320P	80020.001	10/21/1998	60/105,142
2750-0321P	80021.001	10/22/1998	60/105,533
2750-0322P	80022.001	10/26/1998	60/105,571
2750-0323P	80023.001	10/27/1998	60/105,815
2750-0324P	80024.001	10/29/1998	60/106,105
2750-0325P	80025.001	10/30/1998	60/106,218
2750-0326P	80026.001	11/2/1998	60/106,685
2750-0327P	80027.001	11/6/1998	60/107,282
2750-0329P	80029.001	11/9/1998	60/107,719
2750-0328P	80028.001	11/9/1998	60/107,720
2750-0330P	80030.001	11/10/1998	60/107,836
2750-0331P	80031.001	11/12/1998	60/108,190
2750-0332P	80032.001	11/16/1998	60/108,526
2750-0333P	80033.001	11/17/1998	60/108,901
2750-0335P	80035.001	11/19/1998	60/109,127
2750-0334P	80034.001	11/19/1998	60/109,124
2750-0336P	80036.001	11/20/1998	60/109,267
2750-0337P	80037.001	11/23/1998	60/109,594
2750-0338P	80038.001	11/25/1998	60/110,053
2750-0339P	80039.001	11/25/1998	60/110,050
2750-0340P	80040.001	11/27/1998	60/110,158
2750-0341P	80041.001	11/30/1998	60/110,263
2750-0342P	80042.001	12/1/1998	60/110,495
2750-0343P	80043.001	12/2/1998	60/110,626
2750-0344P	80044.001	12/3/1998	60/110,701
2750-0345P	80045.001	12/7/1998	60/111,339
2750-0346P	80046.001	12/9/1998	60/111,589
2750-0347P	80047.001	12/10/1998	60/111,782
2750-0348P	80048.001	12/11/1998	60/111,812
2750-0349P	80049.001	12/14/1998	60/112,096
2750-0350P	80050.001	12/15/1998	60/112,224
2750-0351P	80051.001	12/16/1998	60/112,624
2750-0352P	80052.001	12/17/1998	60/112,862
2750-0353P	80053.001	12/18/1998	60/112,912
2750-0354P	80054.001	12/21/1998	60/113,248

Attorney Docket	Client Docket	Filing Date	Application
2750-0355P	80055.001	12/22/1998	60/113,522
2750-0356P	80056.001	12/23/1998	60/113,826
2750-0357P	80057.001	12/28/1998	60/113,998
2750-0358P	80058.001	12/29/1998	60/114,384
2750-0359P	80059.001	12/30/1998	60/114,455
2750-0360P	80060.001	1/4/1999	60/114,740
2750-0361P	80061.001	1/6/1999	60/114,866
2750-0362P	80062.001	1/7/1999	60/115,153
2750-0367P	80067.001	1/7/1999	60/115,154
2750-0366P	80066.001	1/7/1999	60/115,156
2750-0365P	80065.001	1/7/1999	60/115,155
2750-0363P	80063.001	1/7/1999	60/115,152
2750-0364P	80064.001	1/7/1999	60/115,151
2750-0370P	80070.001	1/8/1999	60/115,293
2750-0369P	80069.001	1/8/1999	60/115,365
2750-0368P	80068.001	1/8/1999	60/115,364
2750-0371P	80071.001	1/11/1999	60/115,339
2750-0372P	80072.001	1/12/1999	60/115,518
2750-0373P	80073.001	1/13/1999	60/115,847
2750-0374P	80074.001	1/14/1999	60/115,905
2750-0375P	80075.001	1/15/1999	60/116,383
2750-0376P	80076.001	1/15/1999	60/116,384
2750-0378P	80078.001	1/19/1999	60/116,340
2750-0377P	80077.001	1/19/1999	60/116,329
2750-0380P	80080.001	1/21/1999	60/116,672
2750-0379P	80079.001	1/21/1999	60/116,674
2750-0381P	80081.001	1/22/1999	60/116,960
2750-0382P	80082.001	1/22/1999	60/116,962
2750-0383P	80083.001	1/28/1999	60/117,756
2750-0384P	80084.001	2/3/1999	60/118,672
2750-0385P	80085.001	2/4/1999	60/118,808
2750-0386P	80086.001	2/5/1999	60/118,778
2750-0387P	80087.001	2/8/1999	60/119,029
2750-0388P	80088.001	2/9/1999	60/119,332
2750-0389P	80089.001	2/10/1999	60/119,462
2750-0391P	80091.001	2/12/1999	60/119,922
2750-0392P	80092.001	2/16/1999	60/120,196
2750-0393P	80093.001	2/16/1999	60/120,198
2750-0394P	80094.001	2/18/1999	60/120,583
2750-0395P	80095.001	2/22/1999	60/121,072
2750-0396P	80096.001	2/23/1999	60/121,334
2750-0397P	80097.001	2/24/1999	60/121,470
2750-0390P	80090.001	2/25/1999	60/121,825
2750-0398P	80098.001	2/25/1999	60/121,704
2750-0399P	80099.001	2/26/1999	60/122,107
2750-0400P	80100.001	3/1/1999	60/122,266
2750-0401P	80101.001	3/2/1999	60/122,568
2750-0402P	80102.001	3/3/1999	60/122,611
2750-0403P	80103.001	3/4/1999	60/121,775

2025 RELEASE UNDER E.O. 14176

Attorney Docket	Client Docket	Filing Date	Application
2750-0405P	80105.001	3/5/1999	60/123,180
2750-0404P	80104.001	3/5/1999	60/123,534
2750-0406P	80106.001	3/9/1999	60/123,680
2750-0407P	80107.001	3/9/1999	60/123,548
2750-0408P	80108.001	3/10/1999	60/123,715
2750-0409P	80109.001	3/10/1999	60/123,726
2750-0410P	80110.001	3/11/1999	60/124,263
2750-0411P	80111.001	3/12/1999	60/123,941
2750-0412P	80112.001	3/23/1999	60/125,788
2750-0413P	80113.001	3/25/1999	60/126,264
2750-0414P	80114.001	3/29/1999	60/126,785
2750-0415P	80115.001	4/1/1999	60/127,462
2750-0416P	91000.001	4/6/1999	60/128,234
2750-0417P	91001.001	4/8/1999	60/128,714
2750-0418P	80118.001	4/16/1999	60/129,845
2750-0420P	80120.001	4/19/1999	60/130,077
2750-0421P	80121.001	4/21/1999	60/130,449
2750-0303P	80115.002	4/23/1999	60/130,510
2750-0422P	80122.001	4/23/1999	60/130,891
2750-0423P	80123.001	4/28/1999	60/131,449
2750-0424P	80124.001	4/30/1999	60/132,407
2750-0425P	80125.001	4/30/1999	60/132,048
2750-0426P	80126.001	5/4/1999	60/132,484
2750-0427P	80127.001	5/5/1999	60/132,485
2750-0428P	91002.001	5/6/1999	60/132,487
2750-0429P	80129.001	5/6/1999	60/132,486
2750-0430P	80130.001	5/7/1999	60/132,863
2750-0431P	80131.001	5/11/1999	60/134,256
2750-0433P	00025.001	5/14/1999	60/134,221
2750-0432P	91006.001	5/14/1999	60/134,370
2750-0434P	80116.001	5/14/1999	60/134,219
2750-0435P	80117.001	5/14/1999	60/134,218
2750-0436P	91007.001	5/18/1999	60/134,768
2750-0437P	91008.001	5/19/1999	60/134,941
2750-0438P	91009.001	5/20/1999	60/135,124
2750-0439P	91010.001	5/21/1999	60/135,353
2750-0440P	91011.001	5/24/1999	60/135,629
2750-0441P	91012.001	5/25/1999	60/136,021
2750-0442P	91013.001	5/27/1999	60/136,392
2750-0444P	91014.001	5/28/1999	60/136,782
2750-0445P	91015.001	6/1/1999	60/137,222
2750-0446P	91016.001	6/3/1999	60/137,528
2750-0447P	91017.001	6/4/1999	60/137,502
2750-0449P	91018.001	6/7/1999	60/137,724
2750-0450P	91019.001	6/8/1999	60/138,094
2750-0457P	00033.001	6/10/1999	60/138,540
2750-0458P	00033.002	6/10/1999	60/138,847
2750-0463P	00034.001	6/14/1999	60/139,119
2750-0461P	80132.011	6/16/1999	60/139,453

20040209 09:22:50

Attorney Docket	Client Docket	Filing Date	Application
2750-0462P	80132.012	6/16/1999	60/139,452
2750-0464P	00037.001	6/17/1999	60/139,492
2750-0453P	80132.005	6/18/1999	60/139,462
2750-0466P	00039.001	6/18/1999	60/139,750
2750-0465P	00038.001	6/18/1999	60/139,763
2750-0460P	80132.010	6/18/1999	60/139,455
2750-0451P	80132.003	6/18/1999	60/139,459
2750-0454P	80132.006	6/18/1999	60/139,457
2750-0459P	80132.009	6/18/1999	60/139,463
2750-0448P	80132.002	6/18/1999	60/139,454
2750-0443P	80132.001	6/18/1999	60/139,458
2750-0456P	80132.008	6/18/1999	60/139,456
2750-0455P	80132.007	6/18/1999	60/139,460
2750-0452P	80132.004	6/18/1999	60/139,461
2750-0467P	00042.001	6/21/1999	60/139,817
2750-0468P	00043.001	6/22/1999	60/139,899
2750-0470P	00042.002	6/23/1999	60/140,353
2750-0469P	00044.001	6/23/1999	60/140,354
2750-0471P	00045.001	6/24/1999	60/140,695
2750-0472P	00046.001	6/28/1999	60/140,823
2750-0473P	00048.001	6/29/1999	60/140,991
2750-0474P	00049.001	6/30/1999	60/141,287
2750-0475P	00050.001	7/1/1999	60/141,842
2750-0476P	00051.001	7/1/1999	60/142,154
2750-0477P	00052.001	7/2/1999	60/142,055
2750-0478P	00053.001	7/6/1999	60/142,390
2750-0479P	00054.001	7/8/1999	60/142,803
2750-0480P	00058.001	7/9/1999	60/142,920
2750-0481P	00059.001	7/12/1999	60/142,977
2750-0482P	00060.001	7/13/1999	60/143,542
2750-0489P	00061.001	7/14/1999	60/143,624
2750-0490P	00062.001	7/15/1999	60/144,005
2750-0485P	80134.003	7/16/1999	60/144,086
2750-0486P	80134.004	7/16/1999	60/144,085
2750-0497P	00064.001	7/19/1999	60/144,325
2750-0496P	80134.014	7/19/1999	60/144,334
2750-0495P	80134.013	7/19/1999	60/144,335
2750-0494P	80134.010	7/19/1999	60/144,333
2750-0492P	80134.008	7/19/1999	60/144,331
2750-0488P	80134.006	7/19/1999	60/144,332
2750-0500P	00065.001	7/20/1999	60/144,632
2750-0502P	80135.002	7/20/1999	60/144,884
2750-0499P	80134.012	7/20/1999	60/144,352
2750-0503P	00066.001	7/21/1999	60/144,814
2750-0483P	80134.001	7/21/1999	60/145,088
2750-0484P	80134.002	7/21/1999	60/145,086
2750-0504P	00067.001	7/22/1999	60/145,192
2750-0491P	80134.007	7/22/1999	60/145,085
2750-0493P	80134.009	7/22/1999	60/145,087

00505226-061000

Attorney Docket	Client Docket	Filing Date	Application
2750-0487P	80134.005	7/22/1999	60/145,089
2750-0498P	80134.011	7/23/1999	60/145,145
2750-0501P	80135.001	7/23/1999	60/145,224
2750-0505P	00069.001	7/23/1999	60/145,218
2750-0506P	00070.001	7/26/1999	60/145,276
2750-0507P	80136.001	7/27/1999	60/145,918
2750-0508P	80136.002	7/27/1999	60/145,919
2750-0509P	00071.001	7/27/1999	60/145,913
2750-0510P	00072.001	7/28/1999	60/145,951
2750-0513P	00073.001	8/2/1999	60/146,386
2750-0512P	80137.002	8/2/1999	60/146,389
2750-0511P	80137.001	8/2/1999	60/146,388
2750-0514P	00074.001	8/3/1999	60/147,038
2750-0515P	00076.001	8/4/1999	60/147,204
2750-0517P	80138.002	8/4/1999	60/147,302
2750-0519P	80136.003	8/5/1999	60/147,192
2750-0518P	00077.001	8/5/1999	60/147,260
2750-0520P	00079.001	8/6/1999	60/147,416
2750-0516P	80138.001	8/6/1999	60/147,303
2750-0523P	80139.002	8/9/1999	60/147,935
2750-0521P	00080.001	8/9/1999	60/147,493
2750-0522P	80139.001	8/10/1999	60/148,171
2750-0524P	00081.001	8/11/1999	60/148,319
2750-0530P	00082.001	8/12/1999	60/148,341
2750-0525P	80141.001	8/12/1999	60/148,347
2750-0526P	80141.002	8/12/1999	60/148,342
2750-0527P	80141.003	8/12/1999	60/148,340
2750-0528P	80141.004	8/12/1999	60/148,337
2750-0532P	80142.002	8/13/1999	60/148,684
2750-0529P	00083.001	8/13/1999	60/148,565
2750-0531P	80142.001	8/16/1999	60/149,368
2750-0533P	80001.002	8/17/1999	60/149,927
2750-0534P	80001.003	8/17/1999	60/149,928
2750-0535P	80001.004	8/17/1999	60/149,926
2750-0536P	80001.005	8/17/1999	60/149,925
2750-0537P	00084.001	8/17/1999	60/149,175
2750-0538P	00085.001	8/18/1999	60/149,426
2750-0542P	00087.001	8/20/1999	60/149,723
2750-0541P	80143.002	8/20/1999	60/149,929
2750-0539P	00086.001	8/20/1999	60/149,722
2750-0543P	00088.001	8/23/1999	60/149,902
2750-0540P	80143.001	8/23/1999	60/149,930
2750-0544P	00089.001	8/25/1999	60/150,566
2750-0547P	00090.001	8/26/1999	60/150,884
2750-0546P	80144.002	8/27/1999	60/151,066
2750-0548P	00091.001	8/27/1999	60/151,080
2750-0545P	80144.001	8/27/1999	60/151,065
2750-0549P	00092.001	8/30/1999	60/151,303
2750-0552P	00093.001	8/31/1999	60/151,438

00555325-061600

U.S. DEPARTMENT OF AGRICULTURE

Attorney Docket	Client Docket	Filing Date	Application
2750-0600P	80026.002	10/28/1999	09/428,944
2750-0601P	00123.001	10/28/1999	60/161,920
2750-0602P	80154.001	10/28/1999	60/161,992
2750-0603P	80154.002	10/28/1999	60/161,993
2750-0604P	00124.001	10/29/1999	60/162,143
2750-0605P	80155.001	10/29/1999	60/162,142
2750-0606P	80155.002	10/29/1999	60/162,228
2750-0607P	00125.001	11/1/1999	60/162,894
2750-0608P	80156.001	11/1/1999	60/162,891
2750-0609P	80156.002	11/1/1999	60/162,895
2750-0610P	00126.001	11/2/1999	60/163,093
2750-0611P	80157.001	11/2/1999	60/163,092
2750-0612P	80157.002	11/2/1999	60/163,091
2750-0614P	80158.001	11/3/1999	60/163,248
2750-0615P	80158.002	11/3/1999	60/163,281
2750-0613P	00127.001	11/3/1999	60/163,249
2750-0618P	80159.002	11/4/1999	60/163,380
2750-0617P	80159.001	11/4/1999	60/163,381
2750-0616P	00128.001	11/4/1999	60/163,379
2750-0621P	80160.002	11/8/1999	60/164,150
2750-0620P	80160.001	11/8/1999	60/164,151
2750-0619P	00129.001	11/8/1999	60/164,146
2750-0623P	80161.002	11/9/1999	60/164,260
2750-0625P	80162.002	11/9/1999	60/164,259
2750-0626P	80163.001	11/10/1999	60/164,321
2750-0630P	80164.002	11/10/1999	60/164,548
2750-0629P	80164.001	11/10/1999	60/164,545
2750-0627P	80163.002	11/10/1999	60/164,318
2750-0624P	80162.001	11/10/1999	60/164,317
2750-0622P	80161.001	11/10/1999	60/164,319
2750-0628P	00131.001	11/10/1999	60/164,544
2750-0636P	80166.002	11/12/1999	60/164,962
2750-0633P	80165.002	11/12/1999	60/164,960
2750-0634P	00133.001	11/12/1999	60/164,870
2750-0632P	80165.001	11/12/1999	60/164,871
2750-0631P	00132.001	11/12/1999	60/164,961
2750-0635P	80166.001	11/12/1999	60/164,959
2750-0637P	00134.001	11/15/1999	60/164,927
2750-0638P	80167.001	11/15/1999	60/164,929
2750-0639P	80167.002	11/15/1999	60/164,926
2750-0640P	00135.001	11/16/1999	60/165,669
2750-0642P	80168.002	11/16/1999	60/165,661
2750-0641P	80168.001	11/16/1999	60/165,671
2750-0643P	00136.001	11/17/1999	60/165,919
2750-0644P	80169.001	11/17/1999	60/165,918
2750-0645P	80169.002	11/17/1999	60/165,911
2750-0646P	00137.001	11/18/1999	60/166,157
2750-0647P	80170.001	11/18/1999	60/166,173
2750-0648P	80170.002	11/18/1999	60/166,158

2025-06-10 09:53:22

Attorney Docket	Client Docket	Filing Date	Application
2750-0650P	80171.001	11/19/1999	60/166,411
2750-0649P	00139.001	11/19/1999	60/166,419
2750-0651P	80171.002	11/19/1999	60/166,412
2750-0653P	80172.001	11/22/1999	60/166,750
2750-0652P	00140.001	11/22/1999	60/166,733
2750-0655P	80173.002	11/23/1999	60/167,362
2750-0654P	80173.001	11/24/1999	60/167,382
2750-0657P	80174.001	11/24/1999	60/167,234
2750-0656P	00141.001	11/24/1999	60/167,233
2750-0658P	80174.002	11/24/1999	60/167,235
2750-0660P	80175.001	11/30/1999	60/167,908
2750-0659P	00142.001	11/30/1999	60/167,904
2750-0661P	80175.002	11/30/1999	60/167,902
2750-0664P	80176.001	12/1/1999	60/168,233
2750-0662P	80042.002	12/1/1999	09/451,320
2750-0665P	80176.002	12/1/1999	60/168,231
2750-0663P	00143.001	12/1/1999	60/168,232
2750-0668P	80177.002	12/2/1999	60/168,548
2750-0667P	80177.001	12/2/1999	60/168,549
2750-0666P	00144.001	12/2/1999	60/168,546
2750-0669P	00145.001	12/3/1999	60/168,675
2750-0670P	80178.001	12/3/1999	60/168,673
2750-0671P	80178.002	12/3/1999	60/168,674
2750-0673P	80179.001	12/7/1999	60/169,278
2750-0672P	00147.001	12/7/1999	60/169,298
2750-0674P	80179.002	12/7/1999	60/169,302
2750-0675P	80180.001	12/8/1999	60/169,692
2750-0676P	80180.002	12/8/1999	60/169,691
2750-0677P	00149.001	12/16/1999	60/171,107
2750-0678P	80181.001	12/16/1999	60/171,114
2750-0679P	80181.002	12/16/1999	60/171,098
2750-0683P	80060.002	1/4/2000	09/478,081
2750-0686F(PC)	80070.100	1/7/2000	00/004,66
2750-0684P	80070.002	1/7/2000	09/479,221
2750-0685P	80183.002	1/19/2000	60/176,867
2750-0688P	80184.002	1/19/2000	60/176,910
2750-0681P	80182.002	1/19/2000	60/176,866
2750-0689P	00152.001	1/26/2000	60/178,166
2750-0691P	80185.001	1/27/2000	60/177,666
2750-0687P	80184.001	1/27/2000	60/178,545
2750-0682P	80183.001	1/27/2000	60/178,546
2750-0680P	80182.001	1/27/2000	60/178,544
2750-0690P	00153.001	1/27/2000	60/178,547
2750-0692P	00155.001	1/28/2000	60/178,754
2750-0693P	80186.001	1/28/2000	60/178,755
2750-0695P	00157.001	2/1/2000	60/179,395
2750-0696P	80187.001	2/1/2000	60/179,388
2750-0694P	80084.002	2/3/2000	09/497,191
2750-0697P	00158.001	2/3/2000	60/180,039

00503726-001600

Attorney Docket	Client Docket	Filing Date	Application
2750-0783F(EP)	91000.101	4/6/2000	00/302,919
2750-0793P	00193.001	4/6/2000	
2750-0784P	91045.001	4/6/2000	
2750-0795P	00194.001	4/6/2000	
2750-0794P	80224.001	4/6/2000	
2750-0783F(MX)	91000.100	4/6/2000	00/003,391
2750-0799P	80226.001	4/7/2000	
2750-0797P	91046.001	4/7/2000	
2750-0798P	00195.001	4/7/2000	60/195,283
2750-0804P	80228.001	4/11/2000	
2750-0801P	80227.002	4/11/2000	60/196,211
2750-0802P	91047.001	4/11/2000	60/196,168
2750-0803P	00196.001	4/11/2000	
2750-0800P	80227.001	4/12/2000	60/196,212
2750-0805P	91048.001	4/12/2000	60/196,483
2750-0806P	00197.001	4/12/2000	60/196,487
2750-0807P	80229.001	4/12/2000	
2750-0809P	80230.001	4/12/2000	60/196,486
2750-0808P	00200.001	4/12/2000	60/196,485
2750-0811P	80231.002	4/13/2000	
2750-0814P	91049.001	4/14/2000	60/197,397
2750-0810P	80231.001	4/14/2000	
2750-0813P	80232.002	4/17/2000	60/197,871
2750-0812P	80232.001	4/17/2000	60/197,870
2750-0817P	91050.001	4/17/2000	60/198,268
2750-0816P	80233.001	4/17/2000	60/197,678
2750-0815P	00201.001	4/17/2000	60/197,687
2750-0819P	80234.001	4/17/2000	60/197,671
2750-0818P	00202.001	4/17/2000	60/198,133
2750-0820P	91051.001	4/19/2000	60/198,400
2750-0821P	00203.001	4/19/2000	60/198,386
2750-0822P	80235.001	4/19/2000	60/198,373
2750-0823P	91052.001	4/20/2000	60/198,629
2750-0824P	00204.001	4/20/2000	60/198,619
2750-0825P	80236.001	4/20/2000	60/198,623
2750-0828P	80237.001	4/21/2000	60/198,763
2750-0826P	91053.001	4/21/2000	
2750-0827P	00206.001	4/21/2000	60/198,767
2750-0829P	91054.001	4/24/2000	
2750-0830P	00207.001	4/24/2000	
2750-0831P	80238.001	4/24/2000	
2750-0833P	92002.001	4/26/2000	
2750-0834P	00208.001	4/26/2000	
2750-0835P	80239.001	4/26/2000	
2750-0832P	92001.001	4/26/2000	60/200,034
2750-0837P	80240.001	4/27/2000	
2750-0836P	00210.001	4/27/2000	
2750-0844P	80242.002	4/28/2000	
2750-0846P	80243.002	4/28/2000	

Attorney Docket	Client Docket	Filing Date	Application
2750-0788P	80123.002	4/28/2000	
2750-0848P	80244.002	5/1/2000	
2750-0839P	80241.001	5/1/2000	
2750-0845P	80243.001	5/1/2000	
2750-0847P	80244.001	5/1/2000	
2750-0840P	91055.001	5/1/2000	
2750-0843P	80242.001	5/1/2000	
2750-0842P	92002.002	5/1/2000	
2750-0841P	92001.002	5/1/2000	
2750-0838P	00211.001	5/2/2000	
2750-0850P	80245.001	5/2/2000	
2750-0849P	91056.001	5/2/2000	
2750-0852P	80126.002	5/4/2000	
2750-0858P	80246.001	5/4/2000	
2750-0857P	00212.001	5/4/2000	
2750-0856P	91057.001	5/4/2000	
2750-0855P	80130.002	5/5/2000	
2750-0861P	80247.001	5/5/2000	
2750-0859P	91058.001	5/5/2000	
2750-0851F(MX)	91002.102	5/5/2000	
2750-0860P	00213.001	5/5/2000	
2750-0851F(EP)	91002.101	5/5/2000	
2750-0853P	80127.002	5/5/2000	
2750-0854P	80129.002	5/5/2000	
2750-0851F(CA)	91002.100	5/5/2000	
2750-0851P	91002.002	5/5/2000	
2750-0865P	00215.001	5/9/2000	
2750-0866P	80249.001	5/9/2000	
2750-0862P	00214.001	5/9/2000	
2750-0863P	80248.001	5/9/2000	
2750-0864P	91059.001	5/9/2000	
2750-0879P	80252.001	5/10/2000	
2750-0877P	91060.001	5/10/2000	
2750-0878P	00216.001	5/10/2000	
2750-0869P	80251.001	5/11/2000	
2750-0870P	80251.002	5/11/2000	
2750-0867P	80250.001	5/11/2000	
2750-0868P	80250.002	5/11/2000	
2750-0871P	80131.002	5/11/2000	
2750-0880P	91061.001	5/11/2000	
2750-0882P	80253.001	5/11/2000	
2750-0881P	00217.001	5/11/2000	
2750-0875P	91006.002	5/12/2000	
2750-0875F(CA)	91006.100	5/12/2000	
2750-0875F(EP)	91006.101	5/12/2000	
2750-0875F(MX)	91006.102	5/12/2000	
2750-0874P	80116.002	5/12/2000	
2750-0872P	80117.002	5/12/2000	
2750-0883P	91062.001	5/12/2000	

0555326.061600

Attorney Docket	Client Docket	Filing Date	Application
2750-0885P	80254.001	5/12/2000	
2750-0884P	00219.001	5/12/2000	
2750-0873P	00025.002	5/12/2000	
2750-0887P	00220.001	5/15/2000	
2750-0888P	80255.001	5/15/2000	
2750-0886P	91063.001	5/15/2000	
2750-0891P	00221.001	5/16/2000	
2750-0892P	80256.001	5/16/2000	
2750-0889P	92001.003	5/17/2000	
2750-0893P	00222.001	5/17/2000	
2750-0894P	80257.001	5/17/2000	
2750-0890P	92002.003	5/17/2000	
2750-0895P	00223.001	5/18/2000	
2750-0876F(MX)	91007.102	5/18/2000	
2750-0876F(EP)	91007.101	5/18/2000	
2750-0876F(CA)	91007.100	5/18/2000	
2750-0896P	80258.001	5/18/2000	
2750-0876P	91007.002	5/18/2000	
2750-0897P	00224.001	5/19/2000	
2750-0898P	80259.001	5/19/2000	
2750-0901P	80260.001	5/22/2000	
2750-0900P	00225.001	5/22/2000	
2750-0899P	91064.001	5/22/2000	
2750-0903P	80261.001	5/23/2000	
2750-0902P	00226.001	5/23/2000	
2750-0904P	00227.001	5/24/2000	
2750-0905P	80262.001	5/24/2000	
2750-0906P	91065.001	5/25/2000	
2750-0907P	00228.001	5/26/2000	
2750-0911P	80264.001	5/26/2000	
2750-0910P	00229.001	5/26/2000	
2750-0908P	80263.001	5/26/2000	
2750-0909P	91066.001	5/26/2000	
2750-0913P	00230.001	5/30/2000	
2750-0914P	80265.001	5/30/2000	
2750-0912P	91067.001	5/30/2000	
2750-0921P	80268.001	6/1/2000	
2750-0920P	00231.001	6/1/2000	
2750-0919P	91068.001	6/1/2000	
2750-0918P	80267.002	6/1/2000	
2750-0916P	80266.002	6/1/2000	
2750-0915P	80266.001	6/2/2000	
2750-0917P	80267.001	6/2/2000	
2750-0922P	91069.001	6/5/2000	
2750-0923P	00232.001	6/5/2000	
2750-0924P	80269.001	6/5/2000	
2750-0925P	91070.001	6/5/2000	
2750-0926P	00233.001	6/5/2000	
2750-0927P	80270.001	6/5/2000	

09505326.061600

Attorney Docket	Client Docket	Filing Date	Application
2750-0928P	00033.003	6/9/2000	
2750-0929P	91071.001	6/8/2000	
2750-0930P	00234.001	6/8/2000	
2750-0931P	80271.001	6/8/2000	
2750-0932P	00235.001	6/9/2000	
2750-0933P	80272.001	6/9/2000	

All applications listed in the table above are expressly incorporated herein by reference in their entirety and for all purposes.

The SDFs of the invention can also be used as probes to search for genes that are related to the SDF within a species. Such related genes are typically considered to be members of a gene family. In such a case, the sequence similarity will often be concentrated into one or a few fragments of the sequence. The fragments of similar sequence that define the gene family typically encode a fragment of a protein or RNA that has an enzymatic or structural function. The percentage of identity in the amino acid sequence of the domain that defines the gene family is preferably at least 70%, more preferably 80 to 95%, most preferably 85 to 99%. To search for members of a gene family within a species, a low stringency hybridization is usually performed, but this will depend upon the size, distribution and degree of sequence divergence of domains that define the gene family. SDFs encompassing regulatory regions can be used to identify coordinately expressed genes by using the regulatory region sequence of the SDF as a probe.

In the instances where the SDFs are identified as being expressed from genes that confer a particular phenotype, then the SDFs can also be used as probes to assay plants of different species for those phenotypes.

I.C. Methods to Inhibit Gene Expression

The nucleic acid molecules of the present invention can be used to inhibit gene transcription and/or translation. Example of such methods include, without limitation:

Antisense Constructs;
Ribozyme Constructs;
Chimeraplast Constructs;
Co-Suppression;
Transcriptional Silencing; and
Other Methods of Gene Expression.

C.1 Antisense

In some instances it is desirable to suppress expression of an endogenous or
5 exogenous gene. A well-known instance is the FLAVOR-SAVOR™ tomato, in which the
gene encoding ACC synthase is inactivated by an antisense approach, thus delaying softening
of the fruit after ripening. See for example, U.S. Patent No. 5,859,330; U.S. Patent No.
5,723,766; Oeller, et al, *Science*, 254:437-439(1991); and Hamilton et al, *Nature*, 346:284-
287 (1990). Also, timing of flowering can be controlled by suppression of the *FLOWERING*
10 *LOCUS C (FLC)*; high levels of this transcript are associated with late flowering, while
absence of *FLC* is associated with early flowering (S.D. Michaels et al., *Plant Cell* 11:949
(1999). Also, the transition of apical meristem from production of leaves with associated
shoots to flowering is regulated by *TERMINAL FLOWER1*, *APETALA1* and *LEAFY*. Thus,
when it is desired to induce a transition from shoot production to flowering, it is desirable to
15 suppress *TFL1* expression (S.J. Liljegren, *Plant Cell* 11:1007 (1999)). As another instance,
arrested ovule development and female sterility result from suppression of the ethylene
forming enzyme but can be reversed by application of ethylene (D. De Martinis et al., *Plant*
Cell 11:1061 (1999)). The ability to manipulate female fertility of plants is useful in
increasing fruit production and creating hybrids.

20 In the case of polynucleotides used to inhibit expression of an endogenous gene, the
introduced sequence need not be perfectly identical to a sequence of the target endogenous gene.
The introduced polynucleotide sequence will typically be at least substantially identical to the
target endogenous sequence.

Some polynucleotide SDFs in Table 1 represent sequences that are expressed in
25 corn, wheat, rice, soybean *Arabidopsis* and/or other plants. Thus the invention includes using
these sequences to generate antisense constructs to inhibit translation and/or degradation of
transcripts of said SDFs, typically in a plant cell.

To accomplish this, a polynucleotide segment from the desired gene that can hybridize to
the mRNA expressed from the desired gene (the "antisense segment") is operably linked to a
30 promoter such that the antisense strand of RNA will be transcribed when the construct is present
in a host cell. A regulated promoter can be used in the construct to control transcription of the
antisense segment so that transcription occurs only under desired circumstances.

The antisense segment to be introduced generally will be substantially identical to at least a fragment of the endogenous gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. Further, the antisense product may hybridize to the untranslated region instead of or in addition to the coding sequence of the gene. The vectors

5 of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced antisense segment sequence also need not be full length relative to either the primary transcription product or the fully processed mRNA. Generally, a higher percentage of sequence identity can be used to compensate for

10 the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and the full length of the transcript can be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about

15 500 nucleotides is especially preferred.

C.2. Ribozymes

It is also contemplated that gene constructs representing ribozymes and based on the SDFs in TABLE 1 are an object of the invention. Ribozymes can also be used to inhibit

20 expression of genes by suppressing the translation of the mRNA into a polypeptide. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences

25 within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs, which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus

30 (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus,

solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*, 334:585 (1988).

Like the antisense constructs above, the ribozyme sequence fragment necessary for pairing need not be identical to the target nucleotides to be cleaved, nor identical to the sequences in TABLE 1. Ribozymes may be constructed by combining the ribozyme sequence and some fragment of the target gene which would allow recognition of the target gene mRNA by the resulting ribozyme molecule. Generally, the sequence in the ribozyme capable of binding to the target sequence exhibits a percentage of sequence identity with at least 80%, preferably with at least 85%, more preferably with at least 90% and most preferably with at least 95%, even more preferably, with at least 96%, 97%, 98% or 99% sequence identity to some fragment of a sequence in TABLE 1 or the complement thereof. The ribozyme can be equally effective in inhibiting mRNA translation by cleaving either in the untranslated or coding regions. Generally, a higher percentage of sequence identity can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective.

C.3. Chimeraplasts

The SDFs of the invention, such as those described by Table 1, can also be used to construct chimeraplasts that can be introduced into a cell to produce at least one specific nucleotide change in a sequence corresponding to the SDF of the invention. A chimeraplast is an oligonucleotide comprising DNA and/or RNA that specifically hybridizes to a target region in a manner which creates a mismatched base-pair. This mismatched base-pair signals the cell's repair enzyme machinery which acts on the mismatched region resulting in the replacement, insertion or deletion of designated nucleotide(s). The altered sequence is then expressed by the cell's normal cellular mechanisms. Chimeraplasts can be designed to repair mutant genes, modify genes, introduce site-specific mutations, and/or act to interrupt or alter normal gene function (US Pat. Nos. 6,010,907 and 6,004,804; and PCT Pub. No. WO99/58723 and WO99/07865).

C.4. Sense Suppression

The SDFs of Table 1 of the present invention are also useful to modulate gene expression by sense suppression. Sense suppression represents another method of gene

suppression by introducing at least one exogenous copy or fragment of the endogenous sequence to be suppressed.

Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter into the chromosome of a plant or by a self-replicating virus has been shown to be an effective means by which to induce degradation of mRNAs of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184. Inhibition of expression may require some transcription of the introduced sequence.

For sense suppression, the introduced sequence generally will be substantially identical to the endogenous sequence intended to be inactivated. The minimal percentage of sequence identity will typically be greater than about 65%, but a higher percentage of sequence identity might exert a more effective reduction in the level of normal gene products. Sequence identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect would likely apply to any other proteins within a similar family of genes exhibiting homology or substantial homology to the suppressing sequence.

C.5. Transcriptional Silencing

The nucleic acid sequences of the invention, including the SDFs of Table 1, and fragments thereof, contain sequences that can be inserted into the genome of an organism resulting in transcriptional silencing. Such regulatory sequences need not be operatively linked to coding sequences to modulate transcription of a gene. Specifically, a promoter sequence without any other element of a gene can be introduced into a genome to transcriptionally silence an endogenous gene (see, for example, Vaucheret, H et al. (1998) *The Plant Journal* 16: 651-659). As another example, triple helices can be formed using oligonucleotides based on sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto. The oligonucleotide can be delivered to the host cell and can bind to the promoter in the genome to form a triple helix and prevent transcription. An oligonucleotide of interest is one that can bind to the promoter and block binding of a transcription factor to the promoter. In such a case, the oligonucleotide can be complementary to the sequences of the promoter that interact with transcription binding factors.

C.6. Other Methods to Inhibit Gene Expression

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

5 Low frequency homologous recombination can be used to target a polynucleotide insert to a gene by flanking the polynucleotide insert with sequences that are substantially similar to the gene to be disrupted. Sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto can be used for homologous recombination.

10 In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Lechan et al., *Trends in Genetics* **13**:152 (1997). In this method, screening for clones from a library containing random insertions is preferred to identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto. The screening can also be
15 performed by selecting clones or R₁ plants having a desired phenotype.

I.D. Methods of Functional Analysis

The constructs described in the methods under I.C. above can be used to determine the function of the polypeptide encoded by the gene that is targeted by the constructs.

20 Down-regulating the transcription and translation of the targeted gene in the host cell or organisms, such as a plant, may produce phenotypic changes as compared to a wild-type cell or organism. In addition, *in vitro* assays can be used to determine if any biological activity, such as calcium flux, DNA transcription, nucleotide incorporation, etc., are being modulated by the down-regulation of the targeted gene.

25 Coordinated regulation of sets of genes, e.g., those contributing to a desired polygenic trait, is sometimes necessary to obtain a desired phenotype. SDFs of the invention representing transcription activation and DNA binding domains can be assembled into hybrid transcriptional activators. These hybrid transcriptional activators can be used with their corresponding DNA elements (i.e., those bound by the DNA-binding SDFs) to effect coordinated expression of desired genes (J.J. Schwarz et al., *Mol. Cell. Biol.* **12**:266 (1992),
30 A. Martinez et al., *Mol. Gen. Genet.* **261**:546 (1999)).

The SDFs of the invention can also be used in the two-hybrid genetic systems to identify networks of protein-protein interactions (L. McAlister-Henn et al., *Methods* 19:330 (1999), J.C. Hu et al., *Methods* 20:80 (2000), M. Golovkin et al., *J. Biol. Chem.* 274:36428 (1999), K. Ichimura et al., *Biochem. Biophys. Res. Comm.* 253:532 (1998)). The SDFs of the invention can also be used in various expression display methods to identify important protein-DNA interactions (e.g. B. Luo et al., *J. Mol. Biol.* 266:479 (1997)).

I.E. Promoters

The SDFs of the invention are also useful as structural or regulatory sequences in a construct for modulating the expression of the corresponding gene in a plant or other organism, e.g. a symbiotic bacterium. For example, promoter sequences associated to SDFs of Table 1 of the present invention can be useful in directing expression of coding sequences either as constitutive promoters or to direct expression in particular cell types, tissues, or organs or in response to environmental stimuli.

With respect to the SDFs of the present invention a promoter is likely to be a relatively small portion of a genomic DNA (gDNA) sequence located in the first 2000 nucleotides upstream from an initial exon identified in a gDNA sequence or initial "ATG" or methionine codon or translational start site in a corresponding cDNA sequence. Such promoters are more likely to be found in the first 1000 nucleotides upstream of an initial ATG or methionine codon or translational start site of a cDNA sequence corresponding to a gDNA sequence. In particular, the promoter is usually located upstream of the transcription start site. The fragments of a particular gDNA sequence that function as elements of a promoter in a plant cell will preferably be found to hybridize to gDNA sequences presented and described in Table 1 at medium or high stringency, relevant to the length of the probe and its base composition.

Promoters are generally modular in nature. Promoters can consist of a basal promoter that functions as a site for assembly of a transcription complex comprising an RNA polymerase, for example RNA polymerase II. A typical transcription complex will include additional factors such as TF_{II}B, TF_{II}D, and TF_{II}E. Of these, TF_{II}D appears to be the only one to bind DNA directly. The promoter might also contain one or more enhancers and/or suppressors that function as binding sites for additional transcription factors that have the function of modulating

the level of transcription with respect to tissue specificity and of transcriptional responses to particular environmental or nutritional factors, and the like.

Short DNA sequences representing binding sites for proteins can be separated from each other by intervening sequences of varying length. For example, within a particular functional module, protein binding sites may be constituted by regions of 5 to 60, preferably 10 to 30, more preferably 10 to 20 nucleotides. Within such binding sites, there are typically 2 to 6 nucleotides that specifically contact amino acids of the nucleic acid binding protein. The protein binding sites are usually separated from each other by 10 to several hundred nucleotides, typically by 15 to 150 nucleotides, often by 20 to 50 nucleotides. DNA binding sites in promoter elements often display dyad symmetry in their sequence. Often elements binding several different proteins, and/or a plurality of sites that bind the same protein, will be combined in a region of 50 to 1,000 basepairs.

Elements that have transcription regulatory function can be isolated from their corresponding endogenous gene, or the desired sequence can be synthesized, and recombined in constructs to direct expression of a coding region of a gene in a desired tissue-specific, temporal-specific or other desired manner of inducibility or suppression. When hybridizations are performed to identify or isolate elements of a promoter by hybridization to the long sequences presented in TABLE 1, conditions are adjusted to account for the above-described nature of promoters. For example short probes, constituting the element sought, are preferably used under low temperature and/or high salt conditions. When long probes, which might include several promoter elements are used, low to medium stringency conditions are preferred when hybridizing to promoters across species.

If a nucleotide sequence of an SDF, or part of the SDF, functions as a promoter or fragment of a promoter, then nucleotide substitutions, insertions or deletions that do not substantially affect the binding of relevant DNA binding proteins would be considered equivalent to the exemplified nucleotide sequence. It is envisioned that there are instances where it is desirable to decrease the binding of relevant DNA binding proteins to silence or down-regulate a promoter, or conversely to increase the binding of relevant DNA binding proteins to enhance or up-regulate a promoter and vice versa. In such instances, polynucleotides representing changes to the nucleotide sequence of the DNA-protein contact region by insertion of additional nucleotides, changes to identity of relevant nucleotides, including use of chemically-modified bases, or deletion of one or more nucleotides are considered encompassed by the present invention. In addition, fragments of the promoter

sequences described by Table 1 and variants thereof can be fused with other promoters or fragments to facilitate transcription and/or transcription in specific type of cells or under specific conditions.

Promoter function can be assayed by methods known in the art, preferably by measuring activity of a reporter gene operatively linked to the sequence being tested for promoter function. Examples of reporter genes include those encoding luciferase, green fluorescent protein, GUS, neo, cat and bar.

I.F. UTRs and Junctions

Polynucleotides comprising untranslated (UTR) sequences and intron/exon junctions are also within the scope of the invention. UTR sequences include introns and 5' or 3' untranslated regions (5' UTRs or 3' UTRs). Fragments of the sequences shown in TABLE 1 can comprise UTRs and intron/exon junctions.

These fragments of SDFs, especially UTRs, can have regulatory functions related to, for example, translation rate and mRNA stability. Thus, these fragments of SDFs can be isolated for use as elements of gene constructs for regulated production of polynucleotides encoding desired polypeptides.

Introns of genomic DNA segments might also have regulatory functions. Sometimes regulatory elements, especially transcription enhancer or suppressor elements, are found within introns. Also, elements related to stability of heteronuclear RNA and efficiency of splicing and of transport to the cytoplasm for translation can be found in intron elements. Thus, these segments can also find use as elements of expression vectors intended for use to transform plants.

Just as with promoters UTR sequences and intron/exon junctions can vary from those shown in TABLE 1. Such changes from those sequences preferably will not affect the regulatory activity of the UTRs or intron/exon junction sequences on expression, transcription, or translation unless selected to do so. However, in some instances, down- or up-regulation of such activity may be desired to modulate traits or phenotypic or *in vitro* activity.

I.G. Coding Sequences

Isolated polynucleotides of the invention can include coding sequences that encode polypeptides comprising an amino acid sequence encoded by sequences in TABLE 1 or an amino acid sequence presented in TABLE 1.

A nucleotide sequence encodes a polypeptide if a cell (or a cell free *in vitro* system) expressing that nucleotide sequence produces a polypeptide having the recited amino acid sequence when the nucleotide sequence is transcribed and the primary transcript is subsequently processed and translated by a host cell (or a cell free *in vitro* system) harboring the nucleic acid. Thus, an isolated nucleic acid that encodes a particular amino acid sequence can be a genomic sequence comprising exons and introns or a cDNA sequence that represents the product of splicing thereof. An isolated nucleic acid encoding an amino acid sequence also encompasses heteronuclear RNA, which contains sequences that are spliced out during expression, and mRNA, which lacks those sequences.

Coding sequences can be constructed using chemical synthesis techniques or by isolating coding sequences or by modifying such synthesized or isolated coding sequences as described above.

In addition to coding sequences encoding the polypeptide sequences of TABLE 1, which are native to corn, *Arabidopsis*, soybean, rice, wheat, and other plants the isolated polynucleotides can be polynucleotides that encode variants, fragments, and fusions of those native proteins. Such polypeptides are described below in part II.

In variant polynucleotides generally, the number of substitutions, deletions or insertions is preferably less than 20%, more preferably less than 15%; even more preferably less than 10%, 5%, 3% or 1% of the number of nucleotides comprising a particularly exemplified sequence. It is generally expected that non-degenerate nucleotide sequence changes that result in 1 to 10, more preferably 1 to 5 and most preferably 1 to 3 amino acid insertions, deletions or substitutions will not greatly affect the function of an encoded polypeptide. The most preferred embodiments are those wherein 1 to 20, preferably 1 to 10, most preferably 1 to 5 nucleotides are added to, deleted from and/or substituted in the sequences specifically disclosed in TABLE 1.

Insertions or deletions in polynucleotides intended to be used for encoding a polypeptide preferably preserve the reading frame. This consideration is not so important in instances when the polynucleotide is intended to be used as a hybridization probe.

II. Polypeptides and Proteins

IIA. Native polypeptides and proteins

Polypeptides within the scope of the invention include both native proteins as well as
5 variants, fragments, and fusions thereof. Polypeptides of the invention are those encoded by
any of the six reading frames of sequences shown in TABLE 1, preferably encoded by the
three frames reading in the 5' to 3' direction of the sequences as shown.

Native polypeptides include the proteins encoded by the sequences shown in TABLE
1. Such native polypeptides include those encoded by allelic variants.

10 Polypeptide and protein variants will exhibit at least 75% sequence identity to those
native polypeptides of TABLE 1. More preferably, the polypeptide variants will exhibit at least
85% sequence identity; even more preferably, at least 90% sequence identity; more preferably at
least 95%, 96%, 97%, 98%, or 99% sequence identity. Fragments of polypeptide or fragments
15 of polypeptides will exhibit similar percentages of sequence identity to the relevant fragments
of the native polypeptide. Fusions will exhibit a similar percentage of sequence identity in that
fragment of the fusion represented by the variant of the native peptide.

Furthermore, polypeptide variants will exhibit at least one of the functional properties of
the native protein. Such properties include, without limitation, protein interaction, DNA
interaction, biological activity, immunological activity, receptor binding, signal transduction,
20 transcription activity, growth factor activity, secondary structure, three-dimensional structure,
etc. As to properties related to *in vitro* or *in vivo* activities, the variants preferably exhibit at least
60% of the activity of the native protein; more preferably at least 70%, even more preferably at
least 80%, 85%, 90% or 95% of at least one activity of the native protein.

One type of variant of native polypeptides comprises amino acid substitutions, deletions
25 and/or insertions. Conservative substitutions are preferred to maintain the function or activity of
the polypeptide.

Within the scope of percentage of sequence identity described above, a polypeptide of
the invention may have additional individual amino acids or amino acid sequences inserted into
the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof.

30 Likewise, some of the amino acids or amino acid sequences may be deleted from the
polypeptide.

A.1 Antibodies

Isolated polypeptides can be utilized to produce antibodies. Polypeptides of the invention can generally be used, for example, as antigens for raising antibodies by known techniques. The resulting antibodies are useful as reagents for determining the distribution of the antigen protein within the tissues of a plant or within a cell of a plant. The antibodies are also useful for examining the production level of proteins in various tissues, for example in a wild-type plant or following genetic manipulation of a plant, by methods such as Western blotting.

Antibodies of the present invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the polypeptides of the invention are first used to immunize a suitable animal, such as a mouse, rat, rabbit, or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies as detection reagents. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization.

Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating the blood at 4°C for 2-18 hours. The serum is recovered by centrifugation (e.g., 1,000xg for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the method of Kohler and Milstein, *Nature* 256: 495 (1975), or modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells can be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate, or well, coated with the protein antigen. B-cells producing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (e.g., hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated

by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected Mab-secreting hybridomas are then cultured either *in vitro* (e.g., in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

5 Other methods for sustaining antibody-producing B-cell clones, such as by EBV transformation, are known.

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding
10 partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TNB) to a blue pigment, quantifiable with a spectrophotometer.

A.2 In Vitro Applications of Polypeptides

Some polypeptides of the invention will have enzymatic activities that are useful *in vitro*.
15 For example, the soybean trypsin inhibitor (Kunitz) family is one of the numerous families of proteinase inhibitors. It comprises plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases. Thus, these peptides find *in vitro* use in protein purification protocols and perhaps in therapeutic settings requiring topical application of protease inhibitors.

20 Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (ALAD) catalyzes the second step in the biosynthesis of heme, the condensation of two molecules of 5-aminolevulinate to form porphobilinogen and is also involved in chlorophyll biosynthesis (Kaczor et al. (1994) Plant Physiol. 1-4: 1411-7; Smith (1988) Biochem. J. 249: 423-8; Schneider (1976) Z. naturforsch. [C] 31: 55-63). Thus, ALAD proteins can be used as catalysts in synthesis of
25 heme derivatives. Enzymes of biosynthetic pathways generally can be used as catalysts for *in vitro* synthesis of the compounds representing products of the pathway.

Polypeptides encoded by SDFs of the invention can be engineered to provide purification reagents to identify and purify additional polypeptides that bind to them. This allows one to identify proteins that function as multimers or elucidate signal transduction or
30 metabolic pathways. In the case of DNA binding proteins, the polypeptide can be used in a similar manner to identify the DNA determinants of specific binding (S. Pierrou et al., *Anal.*

Biochem. 229:99 (1995), S. Chusacultanaichai et al., *J. Biol. Chem.* 274:23591 (1999), Q. Lin et al., *J. Biol. Chem.* 272:27274 (1997)).

II.B. POLYPEPTIDE VARIANTS, FRAGMENTS, AND FUSIONS

Generally, variants, fragments, or fusions of the polypeptides encoded by the SDFs of the invention can exhibit at least one of the activities of the identified domains and/or related polypeptides described in Table 1 corresponding to the SDF of interest.

II.B.(1) Variants

A type of variant of the native polypeptides comprises amino acid substitutions. Conservative substitutions, described above (see II.), are preferred to maintain the function or activity of the polypeptide. Such substitutions include conservation of charge, polarity, hydrophobicity, size, etc. For example, one or more amino acid residues within the sequence can be substituted with another amino acid of similar polarity that acts as a functional equivalent, for example providing a hydrogen bond in an enzymatic catalysis. Substitutes for an amino acid within an exemplified sequence are preferably made among the members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Within the scope of percentage of sequence identity described above, a polypeptide of the invention may have additional individual amino acids or amino acid sequences inserted into the polypeptide in the middle thereof and/or at the N-terminal and/or C-terminal ends thereof. Likewise, some of the amino acids or amino acid sequences may be deleted from the polypeptide. Amino acid substitutions may also be made in the sequences; conservative substitutions being preferred.

One preferred class of variants are those that comprise (1) the domain of an encoded polypeptide and/or (2) residues conserved between the encoded polypeptide and related polypeptides. For this class of variants, the encoded polypeptide sequence is changed by insertion, deletion, or substitution at positions flanking the domain and/or conserved residues.

Another class of variants includes those that comprise an encoded polypeptide sequence that is changed in the domain or conserved residues by a conservative substitution.

Yet another class of variants includes those that lack one of the *in vitro* activities, or structural features of the encoded polypeptides. One example is polypeptides or proteins produced from genes comprising dominant negative mutations. Such a variant may comprise an encoded polypeptide sequence with non-conservative changes in a particular domain or group of conserved residues.

II.A.(2) FRAGMENTS

Fragments of particular interest are those that comprise a domain identified for a polypeptide encoded by an SDF of the instant invention and variants thereof. Also, fragments that comprise at least one region of residues conserved between an SDF encoded polypeptide and its related polypeptides are of great interest. Fragments are sometimes useful as polypeptides corresponding to genes comprising dominant negative mutations are.

II.A.(3) FUSIONS

Of interest are chimeras comprising (1) a fragment of the SDF encoded polypeptide or variants thereof of interest and (2) a fragment of a polypeptide comprising the same domain. For example, an AP2 helix encoded by a SDF of the invention fused to second AP2 helix from ANT protein, which comprises two AP2 helices. The present invention also encompasses fusions of SDF encoded polypeptides, variants, or fragments thereof fused with related proteins or fragments thereof.

DEFINITION OF DOMAINS

The polypeptides of the invention may possess identifying domains. In addition, the domains within the SDF encoded polypeptide can be defined by the region that exhibits at least 70% sequence identity with the consensus sequences listed in the detailed description below of each of the domains.

The majority of the protein domain descriptions given below are obtained from Prosite, (<http://www.expasy.ch/prosite/>), and Pfam, (<http://pfam.wustl.edu/browse.shtml>).

1. (AAA) AAA-protein family signature

A large family of ATPases has been described [1 to 5] whose key feature is that they share a conserved region of about 220 amino acids that contains an ATP-binding site. This family is now called AAA, for 'A'TPases 'A'ssociated with diverse cellular 'A'ctivities. The proteins that belong to this family either contain one or two AAA domains. Proteins containing two AAA domains:

- Mammalian and drosophila NSF (N-ethylmaleimide-sensitive fusion protein) and the fungal homolog, SEC18. These proteins are involved in intracellular transport between the endoplasmic reticulum and Golgi, as well as between different Golgi cisternae.
- Mammalian transitional endoplasmic reticulum ATPase (previously known as p97 or VCP) which is involved in the transfer of membranes from the endoplasmic reticulum to the golgi apparatus. This protein forms a ring-shaped homooligomer composed of six subunits. The yeast homolog is CDC48 and it may play a role in spindle pole proliferation.
- Yeast protein PAS1, essential for peroxisome assembly and the related protein PAS1 from *Pichia pastoris*.
- Yeast protein AFG2.
- *Sulfolobus acidocaldarius* protein SAV and *Halobacterium salinarum* cdcH which may be part of a transduction pathway connecting light to cell division.

Proteins containing a single AAA domain:

- *Escherichia coli* and other bacteria ftsH (or hflB) protein. FtsH is an ATP-dependent zinc metalloprotease that seems to degrade the heat-shock sigma-32 factor. It is an integral membrane protein with a large cytoplasmic C-terminal domain that contain both the AAA and the protease domains.
- Yeast protein YME1, a protein important for maintaining the integrity of the mitochondrial compartment. YME1 is also a zinc-dependent protease.
- Yeast protein AFG3 (or YTA10). This protein also seems to contain a AAA domain followed by a zinc-dependent protease domain.

Subunits from the regulatory complex of the 26S proteasome [6] which is involved in the ATP-dependent degradation of ubiquitinated proteins:

- a) Mammalian subunit 4 and homologs in other higher eukaryotes, in yeast (gene YTA5) and fission yeast (gene mts2).

- b) Mammalian subunit 6 (TBP7) and homologs in other higher eukaryotes and in yeast (gene YTA2).
- c) Mammalian subunit 7 (MSS1) and homologs in other higher eukaryotes and in yeast (gene CIM5 or YTA3).
- d) Mammalian subunit 8 (P45) and homologs in other higher eukaryotes and in yeast (SUG1 or CIM3 or TBY1) and fission yeast (gene let1).

Other probable subunits such as human TBP1 which seems to influences HIV gene expression by interacting with the virus tat transactivator protein and yeast YTA1 and YTA6.

- Yeast protein BCS1, a mitochondrial protein essential for the expression of the Rieske iron-sulfur protein.
- Yeast protein MSP1, a protein involved in intramitochondrial sorting of proteins.
- Yeast protein PAS8, and the corresponding proteins PAS5 from *Pichia pastoris* and PAY4 from *Yarrowia lipolytica*.
- Mouse protein SKD1 and its fission yeast homolog (SpAC2G11.06).
- *Caenorhabditis elegans* meiotic spindle formation protein mei-1.
- Yeast protein SAP1.
- Yeast protein YTA7.
- *Mycobacterium leprae* hypothetical protein A2126A.

It is proposed that, in general, the AAA domains in these proteins act as ATP-dependent protein clamps [5]. In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, there is a highly conserved region located in the central part of the domain which was used to develop a signature pattern.

Consensus pattern: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-x-R

- [1] Froehlich K.-U., Fries H.W., Ruediger M., Erdmann R., Botstein D., Mecke D. J. Cell Biol. 114:443-453(1991).
- [2] Erdmann R., Wiebel F.F., Flessau A., Rytka J., Beyer A., Froehlich K.-U., Kunau W.-H. Cell 64:499-510(1991).
- [3] Peters J.-M., Walsh M.J., Franke W.W. EMBO J. 9:1757-1767(1990).
- [4] Kunau W.-H., Beyer A., Goette K., Marzioch M., Saidowsky J., Skaletz-Rorowski A., Wiebel F.F. Biochimie 75:209-224(1993).

[5] Confalonieri F., Duguet M. *BioEssays* 17:639-650(1995). [6] Hilt W., Wolf D.H. *Trends Biochem. Sci.* 21:96-102(1996).

2. ABC Membrane (ABC transporter transmembrane region). This family represents a unit of six transmembrane helices. Many members of the ABC transporter family (ABC_tran) have two such regions. See also descriptions of ABC Tran, below, and ABC2 membrane, above.

3. (ABC Tran) ABC transporters family signature. On the basis of sequence similarities a family of related ATP-binding proteins has been characterized [1 to 5]. These proteins are associated with a variety of distinct biological processes in both prokaryotes and eukaryotes, but a majority of them are involved in active transport of small hydrophilic molecules across the cytoplasmic membrane. All these proteins share a conserved domain of some two hundred amino acid residues, which includes an ATP-binding site. These proteins are collectively known as ABC transporters. Proteins known to belong to this family are listed below (references are only provided for recently determined sequences). In prokaryotes: - Active transport systems components: alkylphosphonate uptake (phnC/phnK/ phnL); arabinose (araG); arginine (artP); dipeptide (dcjAD;dppD/dppF); ferric enterobactin (fepC); ferrichrome (fhuC); galactoside (mglA); glutamine (glnQ); glycerol-3-phosphate (ugpC); glycine betaine/L-proline (proV); glutamate/aspartate (gltL); histidine (hisP); iron(III) (sfuC), iron(III) dicitrate (fecE); lactose (lacK); leucine/isoleucine/valine (braF/braG; livF/livG); maltose (malK); molybdenum (modC); nickel (nikD/ nikE); oligopeptide (amiE/amiF; oppD/oppF); peptide (sapD/sapF); phosphate (pstB); putrescine (potG); ribose (rbsA); spermidine/putrescine (potA); sulfate (cysA); vitamin B12 (btuD). - Hemolysin/leukotoxin export proteins hlyB, cyaB and lktB. - Colicin V export protein cvaB. - Lactococcal export protein lcnC [6]. - Lantibiotic transport proteins nisT (nisin) and spaT (subtilin). - Extracellular proteases B and C export protein prtD. - Alkaline protease secretion protein aprD. - Beta-(1,2)-glucan export proteins chvA and ndvA. - *Haemophilus influenzae* capsule-polysaccharide export protein bexA. - Cytochrome c biogenesis proteins ccmA (also known as cycV and helA). - Polysialic acid transport protein kpsT. - Cell division associated ftsE protein (function unknown). - Copper processing protein nosF from *Pseudomonas stutzeri*. - Nodulation protein nodI from *Rhizobium* (function unknown). - *Escherichia coli* proteins cydC and cydD. - Subunit A of the ABC excision nuclease (gene *uvrA*). -

Erythromycin resistance protein from *Staphylococcus epidermidis* (gene *msrA*). - Tylosin resistance protein from *Streptomyces fradiae* (gene *tlrC*) [7]. - Heterocyst differentiation protein (gene *hetA*) from *Anabaena* PCC 7120. - Protein P29 from *Mycoplasma hyorhinis*, a probable component of a high affinity transport system. - *yhbG*, a putative protein whose gene is linked with *ntrA* in many bacteria such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas putida*, *Rhizobium meliloti* and *Thiobacillus ferrooxidans*. - *Escherichia coli* and related bacteria hypothetical proteins *yabJ*, *yadG*, *yagC*, *ybbA*, *ycjW*, *yddA*, *yehX*, *yejF*, *yheS*, *yhiG*, *yhiH*, *yjcW*, *yjjK*, *yojI*, *yrbF* and *ytfR*. In eukaryotes: - The multidrug transporters (Mdr) (P-glycoprotein), a family of closely related proteins which extrude a wide variety of drugs out of the cell (for a review see [8]). - Cystic fibrosis transmembrane conductance regulator (CFTR), which is most probably involved in the transport of chloride ions. - Antigen peptide transporters 1 (TAP1, PSF1, RING4, HAM-1, mtp1) and 2 (TAP2, PSF2, RING11, HAM-2, mtp2), which are involved in the transport of antigens from the cytoplasm to a membrane-bound compartment for association with MHC class I molecules. - 70 Kd peroxisomal membrane protein (PMP70). - ALDP, a peroxisomal protein involved in X-linked adrenoleukodystrophy [9]. - Sulfonyleurea receptor [10], a putative subunit of the B-cell ATP-sensitive potassium channel. - *Drosophila* proteins white (*w*) and brown (*bw*), which are involved in the import of ommatidium screening pigments. - Fungal elongation factor 3 (EF-3). - Yeast STE6 which is responsible for the export of the α -factor pheromone. - Yeast mitochondrial transporter ATM1. - Yeast MDL1 and MDL2. - Yeast SNQ2. - Yeast sporidesmin resistance protein (gene *PDR5* or *STS1* or *YDR1*). - Fission yeast heavy metal tolerance protein *hmt1*. This protein is probably involved in the transport of metal-bound phytochelatin. - Fission yeast brefeldin A resistance protein (gene *bfr1* or *hba2*). - Fission yeast leptomycin B resistance protein (gene *pmd1*). - *mbpX*, a hypothetical chloroplast protein from Liverwort. - Prestalk-specific protein *tagB* from slime mold. This protein consists of two domains: a N-terminal subtilase catalytic domain and a C-terminal ABC transporter domain. As a signature pattern for this class of proteins, a conserved region which is located between the 'A' and the 'B' motifs of the ATP-binding site was used.

Consensus pattern: [LIVMFYC]-[SA]-[SAPGLVFKQH]-G-[DENQMW]-[KROASPCLIMFW]-[KRQSTAVM]-[KRACLVM]-[LIVMFYPAN]-[PHY]-[LIVMFW]-[SAGCLIVP]-[FYWHP]-{KRHP}-[LIVMFYWSTA] The ATP-binding region is duplicated in *araG*, *mdl*, *msrA*, *rbsA*, *tlrC*, *uvrA*, *yejF*, *Mdr*'s, *CFTR*, *pmd1* and in *EF-3*. In

some of those proteins, the above pattern only detect one of the two copies of the domain. The proteins belonging to this family also contain one or two copies of the ATP-binding motifs 'A' and 'B'.

- [1] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).
- [2] Higgins C.F., Gallagher M.P., Mimmack M.M., Pearce S.R. BioEssays 8:111-116(1988).
- [3] Higgins C.F., Hiles I.D., Salmond G.P.C., Gill D.R., Downie J.A., Evans I.J., Holland I.B., Gray L., Buckels S.D., Bell A.W., Hermodson M.A. Nature 323:448-450(1986).
- [4] Doolittle R.F., Johnson M.S., Husain I., van Houten B., Thomas D.C., Sancar A. Nature 323:451-453(1986).
- [5] Blight M.A., Holland I.B. Mol. Microbiol. 4:873-880(1990).
- [6] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L. Appl. Environ. Microbiol. 58:1952-1961(1992).
- [7] Rosteck P.R. Jr., Reynolds P.A., Hersherberger C.L. Gene 102:27-32(1991).
- [8] Gottesman M.M., Pastan I. J. Biol. Chem. 263:12163-12166(1988).
- [9] Valle D., Gaertner J. Nature 361:682-683(1993).
- [10] Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV, Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J., Nelson D.A. Science 268:423-426(1995).

4. (ACBP)

Acyl-CoA-binding protein signature

Acyl-CoA-binding protein (ACBP) is a small (10 Kd) protein that binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters [1]. ACBP is also known as diazepam binding inhibitor (DBI) or endozepine (EP) because of its ability to displace diazepam from the benzodiazepine (BZD) recognition site located on the GABA type A receptor. It is therefore possible that this protein also acts as a neuropeptide to modulate the action of the GABA receptor [2]. ACBP is a highly conserved protein of about 90 residues that has been so far found in vertebrates, insects and yeast. ACBP is also related to the N-terminal section of a probable transmembrane protein of

unknown function which has been found in mammals. As a signature pattern, the region that corresponds to residues 19 to 37 in mammalian ACBP was selected.

Consensus pattern: P-[STA]-x-[DEN]-x-[LIVMF]-x(2)-[LIVMFY]-Y-[GSTA]-x-[FY]-K- Q-
5 [STA](2)-x-G-

[1] Rose T.M., Schultz E.R., Todaro G.J. Proc. Natl. Acad. Sci. U.S.A. 89:11287-11291(1992).

[2] Costa E., Guidotti A. Life Sci. 49:325-344(1991).

5. (AIRS)

AIR synthase related proteins

15 This family includes Hydrogen expression/formation protein Hype, AIR synthases, FGAM synthase and selenide, water dikinase.

6. (AMP-binding)

20 Putative AMP-binding domain signature

It has been shown [1 to 5] that a number of prokaryotic and eukaryotic enzymes which all probably act via an ATP-dependent covalent binding of AMP to their substrate, share a region of sequence similarity. These enzymes are: - Insects luciferase (luciferin 4-monooxygenase). Luciferase produces light by catalyzing the oxidation of luciferin in presence of ATP and molecular oxygen. - Alpha-aminoadipate reductase from yeast (gene LYS2). This enzyme catalyzes the activation of alpha-aminoadipate by ATP-dependent adenylation and the reduction of activated alpha-aminoadipate by NADPH. - Acetate--CoA ligase (acetyl-CoA synthetase), an enzyme that catalyzes the formation of acetyl-CoA from acetate and CoA. - Long-chain-fatty-acid--CoA ligase, an enzyme that activates long-chain fatty acids for both the synthesis of cellular lipids and their degradation via beta-oxidation. - 4-coumarate--CoA ligase (4CL), a plant enzyme that catalyzes the formation of 4-coumarate-CoA from 4-coumarate and coenzyme A; the branchpoint reactions between

general phenylpropanoid metabolism and pathways leading to various specific end products. -

O-succinylbenzoic acid--CoA ligase (OSB-CoA synthetase) (gene menE) [6], a bacterial enzyme involved in the biosynthesis of menaquinone (vitamin K₂). - 4-Chlorobenzoate--CoA

ligase (EC 6.2.1.-) (4-CBA--CoA ligase) [7], a *Pseudomonas* enzyme involved in the

5 degradation of 4-CBA. - Indoleacetate--lysine ligase (IAA-lysine synthetase) [8], an enzyme from *Pseudomonas syringae* that converts indoleacetate to IAA-lysine. - Bile acid-CoA ligase (gene baiB) from *Eubacterium* strain VPI 12708 [4]. This enzyme catalyzes the ATP-

dependent formation of a variety of C-24 bile acid-CoA. - Crotonobetaine/carnitine-CoA

10 ligase (EC 6.3.2.-) from *Escherichia coli* (gene caiC). - L-(alpha-aminoadipyl)-L-cysteinyl-D-

valine synthetase (ACV synthetase) from various fungi (gene acvA or pcbAB). This enzyme catalyzes the first step in the biosynthesis of penicillin and cephalosporin, the formation of

ACV from the constituent amino acids. The amino acids seem to be activated by adenylation.

It is a protein of around 3700 amino acids that contains three related domains of about 1000

amino acids. - Gramicidin S synthetase I (gene grsA) from *Bacillus brevis*. This enzyme

15 catalyzes the first step in the biosynthesis of the cyclic antibiotic gramicidin S, the ATP-dependent racemization of phenylalanine - Tyrocidine synthetase I (gene tycA) from

Bacillus brevis. The reaction carried out by tycA is identical to that catalyzed by grsA -

Gramicidin S synthetase II (gene grsB) from *Bacillus brevis*. This enzyme is a multifunctional protein that activates and polymerizes proline, valine, ornithine and leucine.

20 GrsB consists of four related domains. - Enterobactin synthetase components E (gene entE) and F (gene entF) from *Escherichia coli*. These two enzymes are involved in the ATP-

dependent activation of respectively 2,3-dihydroxybenzoate and serine during enterobactin (enterochelin) biosynthesis. - Cyclic peptide antibiotic surfactin synthase subunits 1, 2 and 3

from *Bacillus subtilis*. Subunits 1 and 2 contains three related domains while subunit 3 only

25 contains a single domain. - HC-toxin synthetase (gene HTS1) from *Cochliobolus carbonum*.

This enzyme activates the four amino acids (Pro, L-Ala, D-Ala and 2-amino-9,10-epoxi-8-oxodecanoic acid) that make up HC-toxin, a cyclic tetrapeptide. HTS1 consists of four related domains. There are also some proteins, whose exact function is not yet known, but which are,

very probably, also AMP-binding enzymes. These proteins are: - ORA (octapeptide-repeat

30 antigen), a *Plasmodium falciparum* protein whose function is not known but which shows a

high degree of similarity with the above proteins. - AngR, a *Vibrio anguillarum* protein.

AngR is thought to be a transcriptional activator which modulates the anguibactin (an iron-

binding siderophore) biosynthesis gene cluster operon. But it is believed [9], that angR is not

a DNA-binding protein, but rather an enzyme involved in the biosynthesis of anguibactin.

This conclusion is based on three facts: the presence of the AMP-binding domain; the size of angR (1048 residues), which is far bigger than any bacterial transcriptional protein; and the presence of a probable S-acyl thioesterase immediately downstream of angR. - A

5 hypothetical protein in mmsB 3' region in *Pseudomonas aeruginosa*. - *Escherichia coli* hypothetical protein ydiD. - Yeast hypothetical protein YBR041w. - Yeast hypothetical protein YBR222c. - Yeast hypothetical protein YER147c. All these proteins contain a highly conserved region very rich in glycine, serine, and threonine which is followed by a conserved lysine. A parallel can be drawn between this type of domain and the G-x(4)-G-K-[ST] ATP-
10 /GTP-binding 'P-loop' domain or the protein kinases G-x-G-x(2)-[SG]-x(10,20)-KATP-binding domains.

Consensus pattern: [LIVMFY]-x(2)-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR] In a majority of cases the residue that follows the Lys at the end of the pattern is a Gly.

- [1] Toh H. Protein Seq. Data Anal. 4:111-117(1991).
- [2] Smith D.J., Earl A.J., Turner G. EMBO J. 9:2743-2750(1990).
- [3] Schroeder J. Nucleic Acids Res. 17:460-460(1989).
- [4] Mallonee D.H., Adams J.L., Hylemon P.B. J. Bacteriol. 174:2065-2071(1992).
- 20 [5] Turgay K., Krause M., Marahiel M.A. Mol. Microbiol. 6:529-546(1992).
- [6] Driscoll J.R., Taber H.W. J. Bacteriol. 174:5063-5071(1992).
- [7] Babbitt P.C., Kenyon G.L., Matin B.M., Charest H., Sylvestre M., Scholten J.D., Chang K.-H., Liang P.-H., Dunaway-Mariano D. Biochemistry 31:5594-5604(1992).
- [8] Farrell D.H., Mikesell P., Actis L.A., Crosa J.H. Gene 86:45-51(1990).

7. AP2 domain

This 60 amino acid residue domain can bind to DNA [1]. This domain is plant specific.

30 Members of this family are suggested to be related to pyridoxal phosphate-binding domains such as found in aminotran_2 [3]. AP2 domains are also described in Jofuku et al., co-pending U.S. Patent applications 08/700,152, 08/879,827, 08/912,272, 09/026,039.

- [1] Ohme-takagi M, Shinshi H; Plant Cell 1995;7:173-182.
 [2] Weigel D; Plant Cell 1995;7:388-389.
 [3] Mushegian AR, Koonin EV; Genetics 1996;144:817-828.

5

8. ARID

The ARID domain is an AT-Rich Interaction domain sharing structural homology to DNA replication and repair nucleases and polymerases.

- 10 [1] Herscher RF, Kaplan MH, Lelsz DL, Das C, Scheuermann R, Tucker PW; Genes Dev 1995;9:3067-3082.
 [2] Yuan YC, Whitson RH, Liu Q, Itakura K, Chen Y; Nat Struct Biol 1998;5:959-964.

15 9. (ATP synt)

ATP synthase gamma subunit signature

- 20 ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), and a catalytic core, called coupling factor CF(1). The former acts as a proton channel; the latter is composed of five subunits, alpha, beta, gamma, delta and epsilon. Subunit gamma is believed to be important in regulating ATPase activity and the flow of protons through the CF(0) complex. The best conserved region of the gamma subunit [3] is its C-terminus which seems to be essential for assembly and catalysis. As a signature pattern to detect ATPase gamma subunits, a14 residue conserved segment where the last amino acid is found one to three residues from the C-terminal extremity was used.

- 25 Consensus pattern: [IV]-T-x-E-x(2)-[DE]-x(3)-G-A-x-[SAKR]- Note: Pea chloroplast gamma and two Bacillus species gamma subunits are not detected by this motif.

- 30 [1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).
 [2] Senior A.E. Physiol. Rev. 68:177-231(1988).

[3] Miki J., Maeda M., Mukohata Y., Futai M. FEBS Lett. 232:221-226(1988).

10. (ATP Synt A)

5 Synthase a subunit signature

ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1). The CF(0) a subunit, also called protein 6, is a key component of the proton channel; it may play a direct role in translocating protons across the membrane. It is a highly hydrophobic protein that has been predicted to contain 8 transmembrane regions [3]. Sequence comparison of a subunits from all available sources reveals very few conserved regions. The best conserved region is located in what is predicted to be the fifth transmembrane domain. This region contains three perfectly conserved residues: an arginine, a leucine and an asparagine. Mutagenesis experiments of ATPase activity. This region was selected as a signature pattern.

Consensus pattern: [STAGN]-x-[STAG]-[LIVMF]-R-L-x-[SAGV]-N-[LIVMT] [R is important for proton translocation]

[1] Futai M., Noumi T., Maeda M. Annu. Rev. Biochem. 58:111-136(1989).

[2] Senior A.E. Physiol. Rev. 68:177-231(1988).

[3] Lewis M.L., Chang J.A., Simoni R.D. J. Biol. Chem. 265:10541-10550(1990).

[4] Cain B.D., Simoni R.D. J. Biol. Chem. 264:3292-3300(1989).

11. ATP synthase B

Part of the CF(0) (base unit) of the ATP synthase. The base-unit is thought to translocate protons through membrane (inner membrane in mitochondria, thylakoid membrane in plants, cytoplasmic membrane in bacteria). The B subunits are thought to interact with the stalk of the CF(1) subunits.

12. (ATP synt C)

ATP synthase c subunit signature

ATP synthase (proton-translocating ATPase) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1). The CF(0) c subunit (also called protein 9, proteolipid, or subunit III) [3,4] is a highly hydrophobic protein of about 8 Kd which has been implicated in the proton-conducting activity of ATPase. Structurally subunit c consists of two long terminal hydrophobic regions, which probably span the membrane, and a central hydrophilic region. N,N'-dicyclohexylcarbodiimide (DCCD) can bind covalently to subunit c and thereby abolish the ATPase activity. DCCD binds to a specific glutamate or aspartate residue which is located in the middle of the second hydrophobic region near the C-terminus of the protein. A signature pattern which includes the DCCD-binding residue was derived.

Consensus pattern: [GSTA]-R-[NQ]-P-x(10)-[LIVMFYW](2)-x(3)-[LIVMFYW]-x-[DE] [D or E binds DCCD]

[1] Futai M., Noumi T., Maeda M. *Annu. Rev. Biochem.* 58:111-136(1989).

[2] Senior A.E. *Physiol. Rev.* 68:177-231(1988).

[3] Ivaschenko A.T., Karpenyuk T.A., Ponomarenko S.V. *Biokhimiia* 56:406-419(1991).

[4] Recipon H., Perasso R., Adoutte A., Quetier F. J. *Mol. Evol.* 34:292-303(1992).

13. (ATP synt DE)

ATP synthase, Delta/Epsilon chain

Part of the ATP synthase CF(1). These subunits are part of the head unit of the ATP synthase. The subunits are called delta and epsilon in human and metazoan species but in bacterial

species the delta (D) subunit is theequivalent to the Oligomycin sensitive subunit (OSCP) in metozoans.

5 14. (ATP synt ab)

ATP synthase alpha and beta subunits signature

ATP synthase (proton-translocating ATPase) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), and a catalytic core, called coupling factor CF(1). The former acts as a proton channel; the latter is composed of five subunits, alpha, beta, gamma, delta and epsilon. The sequences of subunits alpha and beta are related and both contain a nucleotide-binding site for ATP and ADP. The beta chain has catalytic activity, while the alpha chain is a regulatory subunit. Vacuolar ATPases [3] (V-ATPases) are responsible for acidifying a variety of intracellular compartments in eukaryotic cells. Like F-ATPases, they are oligomeric complexes of a transmembrane and a catalytic sector. The sequence of the largest subunit of the catalytic sector (70 Kd) is related to that of F-ATPase beta subunit, while a 60 Kd subunit, from the same sector, is related to the F-ATPases alpha subunit [4]. Archaeobacterial membrane-associated ATPases are composed of three subunits. The alpha chain is related to F-ATPases beta chain and the beta chain is related to F-ATPases alpha chain [4]. A protein highly similar to F-ATPase beta subunits is found [5] in some bacterial apparatus involved in a specialized protein export pathway that proceeds without signal peptide cleavage. This protein is known as flil in *Bacillus* and *Salmonella*, Spa47 (mxlB) in *Shigella flexneri*, HrpB6 in *Xanthomonas campestris* and yscN in *Yersinia* virulence plasmids. To detect these ATPase subunits, a segment of ten amino-acid residues, containing two conserved serines, as a signature pattern was selected. The first serine seems to be important for catalysis - in the ATPase alpha chain at least - as its mutagenesis causes catalytic impairment.

30 Consensus pattern: P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S [The first S is a putative active site residue]

[1] Futai M., Noumi T., Maeda M. *Annu. Rev. Biochem.* 58:111-136(1989).

[2] Senior A.E. *Physiol. Rev.* 68:177-231(1988).

[3] Nelson N. J. *Bioenerg. Biomembr.* 21:553-571(1989).

[4] Gogarten J.P., Kibak H., Dittrich P., Taiz L., Bowman E.J., Bowman B.J., Manolson M.F., Poole R.J., Date T., Oshima T., Konishi J., Denda K., Yoshida M. *Proc. Natl. Acad.*

5 *Sci. U.S.A.* 86:6661-6665(1989).

[5] Dreyfus G., Williams A.W., Kawagishi I., MacNab R.M. *J. Bacteriol.* 175:3131-3138(1993).

10 15. (ATP synt ab C)

ATP synthase ab C terminal.

Number of members: 190

15 [1] Abrahams JP, Leslie AG, Lutter R, Walker JE; "Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria." *Nature* 1994;370:621-628.

16. (A deaminase)

20 Adenosine and AMP deaminase signature

Adenosine deaminase catalyzes the hydrolytic deamination of adenosine into inosine. AMP deaminase catalyzes the hydrolytic deamination of AMP into IMP. It has been shown [1] that these two types of enzymes share three regions of sequence similarities; these regions are centered on residues which are proposed to play an important role in the catalytic mechanism of these two enzymes. One of these regions, containing two conserved aspartic acid residues that are potential active site residues was selected.

25 Consensus pattern: [SA]-[LIVM]-[NGS]-[STA]-D-D-P [The two D's are putative active site residues]

30 [1] Chang Z., Nygaard P., Chinault A.C., Kellems R.E. *Biochemistry* 30:2273-2280(1991).

17. (Acetyltransf)

Acetyltransferase (GNAT) family.

- 5 This family contains proteins with N-acetyltransferase functions.

[1] Neuwald AF, Landsman D; Trends Biochem Sci 1997;22:154-155.

10 18. (Aconitase C)

Aconitase family signature

Aconitase (aconitate hydratase) (EC 4.2.1.3) [1] is the enzyme from the tricarboxylic acid cycle that catalyzes the reversible isomerization of citrate and isocitrate. Cis-aconitate is formed as an intermediary product during the course of the reaction. In eukaryotes two isoforms of aconitase are known to exist: one found in the mitochondrial matrix and the other found in the cytoplasm. Aconitase, in its active form, contains a 4Fe-4S iron-sulfur cluster; three cysteine residues have been shown to be ligands of the 4Fe-4S cluster. It has been shown that the aconitase family also contains the following proteins: - Iron-responsive element binding protein (IRE-BP). IRE-BP is a cytosolic protein that binds to iron-responsive elements (IREs). IREs are stem-loop structures found in the 5'UTR of ferritin, and delta aminolevulinic acid synthase mRNAs, and in the 3'UTR of transferrin receptor mRNA. IRE-BP also express aconitase activity. - 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate isomerase), the enzyme that catalyzes the second step in the biosynthesis of leucine. - Homoaconitase (EC 4.2.1.36) (homoaconitate hydratase), an enzyme that participates in the alpha-aminoadipate pathway of lysine biosynthesis and that converts cis-homoaconitate into homoisocitric acid. - Escherichia coli protein ybhJ. As a signature for proteins from the aconitase family, two conserved regions that contain the three cysteine ligands of the 4Fe-4S cluster were selected.

Consensus pattern: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-[LIVMA] [C binds the iron-sulfur center]

Consensus pattern: G-x(2)-[LIVWPO]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA] [The two C's bind the iron-sulfur center]

[1] Gruer M.J., Artymiuk P.J., Guest J.R. Trends Biochem. Sci. 22:3-6(1997).

19. (Acyl-CoA dh)

Acyl-CoA dehydrogenases signatures

Acyl-CoA dehydrogenases [1,2,3] are enzymes that catalyze the alpha, beta-dehydrogenation of acyl-CoA esters and transfer electrons to ETF, the electron transfer protein. Acyl-CoA dehydrogenases are FAD flavoproteins. This family currently includes: - Five eukaryotic isozymes that catalyze the first step of the beta-oxidation cycles for fatty acids with various chain lengths. These are short (SCAD) (EC 1.3.99.2), medium (MCAD) (EC 1.3.99.3), long (LCAD) (EC 1.3.99.13), very-long (VLCAD) and short/branched (SBCAD) chain acyl-CoA dehydrogenases. These enzymes are located in the mitochondrion. They are all homotetrameric proteins of about 400 amino acid residues except VLCAD which is a dimer and which contains, in its mature form, about 600 residues. - Glutaryl-CoA dehydrogenase (EC 1.3.99.7) (GCDH), which is involved in the catabolism of lysine, hydroxylysine and tryptophan. - Isovaleryl-CoA dehydrogenase (EC 1.3.99.10) (IVD), involved in the catabolism of leucine. - Acyl-coA dehydrogenases acsA and mmgC from *Bacillus subtilis*. - Butyryl-CoA dehydrogenase (EC 1.3.99.2) from *Clostridium acetobutylicum*. - *Escherichia coli* protein caiA [4]. - *Escherichia coli* protein aidB. Two conserved regions were selected as signature patterns. The first is located in the center of these enzymes, the second in the C-terminal section.

Consensus pattern: [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA]

Consensus pattern: [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)-[DEN]

[1] Tanaka K., Ikeda, Matsubara Y., Hyman D.B. Enzyme 38:91-107(1987).

[2] Matsubara Y., Indo Y., Naito E., Ozasa H., Glassberg R., Vockley J., Ikeda Y., Kraus J., Tanaka K. J. Biol. Chem. 264:16321-16331(1989).

[3] Aoyama T., Ueno I., Kamijo T., Hashimoto T. J. Biol. Chem. 269:19088-19094(1994).

[4] Eichler K., Bourgis F., Buchet A., Kleber H.-P., Mandrand-Berthelot M.-A. Mol. Microbiol. 13:775-786(1994).

5

20. (Acyl transf)

Acyl transferase domain

Number of members: 161

10

[1] Serre L, Verbree EC, Dauter Z, Stuitje AR, Derewenda ZS; Medline: [95286570](#) "The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component." J Biol Chem 1995;270:12961-12964.

15

21. Acylphosphatase signatures

Acylphosphatase (EC [3.6.1.7](#)) [1,2] catalyzes the hydrolysis of various acylphosphate carboxyl-phosphate bonds such as carbamyl phosphate, succinylphosphate, 1,3-diphosphoglycerate, etc. The physiological role of this enzyme is not yet clear.

20

Acylphosphatase is a small protein of around 100 amino-acid residues. There are two known isozymes. One seems to be specific to muscular tissues, the other, called 'organ-common type', is found in many different tissues. While acylphosphatase have been so far only characterized in vertebrates, there are a number of bacterial and archebacterial hypothetical proteins that are highly similar to that enzyme and that probably possess the same activity. These proteins are: - Escherichia coli hypothetical protein yccX. - Bacillus subtilis hypothetical protein yfIL. - Archaeoglobus fulgidus hypothetical protein AF0818. Two conserved regions were selected as signature patterns. The first is located in the N-terminal section, while the second is found in the central part of the protein sequence.

25

Consensus pattern: [LIV]-x-G-x-V-Q-G-V-x-[FM]-R

Consensus pattern: G-[FYW]-[AVC]-[KRQAM]-N-x(3)-G-x-V-x(5)-G

[1] Stefani M., Ramponi G. *Life Chem. Rep.* 12:271-301(1995).

[2] Stefani M., Taddei N., Ramponi G. *Cell. Mol. Life Sci.* 53:141-151(1997).

22. (Adap comp sub)

Clathrin adaptor complexes medium chain signatures.

Clathrin coated vesicles (CCV) mediate intracellular membrane traffic such as receptor mediated endocytosis. In addition to clathrin, the CCV are composed of a number of other components including oligomeric complexes which are known as adaptor or clathrin assembly proteins (AP) complexes [1]. The adaptor complexes are believed to interact with the cytoplasmic tails of membrane proteins, leading to their selection and concentration. In mammals two types of adaptor complexes are known: AP-1 which is associated with the Golgi complex and AP-2 which is associated with the plasma membrane. Both AP-1 and AP-2 are heterotetramers that consist of two large chains - the adaptins - (gamma and beta' in AP-1; alpha and beta in AP-2); a medium chain (AP47 in AP-1; AP50 in AP-2) and a small chain (AP19 in AP-1; AP17 in AP-2). The medium chains of AP-1 and AP-2 are evolutionary related proteins of about 50 Kd. Homologs of AP47 and AP50 have also been found in *Caenorhabditis elegans* (genes *unc-101* and *ap50*) [2] and yeast (gene *APM1* or *YAP54*) [3]. Some more divergent, but clearly evolutionary related proteins have also been found in yeast: *APM2* and *YBR288c*. Two conserved regions were selected as signature patterns, one located in the N-terminal region, the other from the central section of these proteins.

Consensus pattern: [IVT]-[GSP]-W-R-x(2,3)-[GAD]-x(2)-[HY]-x(2)-N-x-[LIVMAFY](3)-D-[LIVM]-[LIVMT]-E

Consensus pattern: [LIV]-x-F-I-P-P-x-G-x-[LIVMFY]-x-L-x(2)-Y

[1] Pearse B.M., Robinson M.S. *Annu. Rev. Cell Biol.* 6:151-171(1990).

[2] Lee J., Jongeward G.D., Sternberg P.W. *Genes Dev.* 8:60-73(1994).

[3] Nakayama Y., Goebel M., O'Brine G.B., Lemmon S., Pingchang C.E., Kirchhausen T. *Eur. J. Biochem.* 202:569-574(1991).

23. (Adenylsucc synt)

Adenylosuccinate synthetase signatures

Adenylosuccinate synthetase (EC 6.3.4.4) [1] plays an important role in purine biosynthesis, by catalyzing the GTP-dependent conversion of IMP and aspartic acid to AMP. Adenylosuccinate synthetase has been characterized from various sources ranging from *Escherichia coli* (gene *purA*) to vertebrate tissues. Invertebrates, two isozymes are present - one involved in purine biosynthesis and the other in the purine nucleotide cycle. Two conserved regions were selected as signature patterns. The first one is a perfectly conserved octapeptide located in the N-terminal section and which is involved in GTP-binding [2]. The second one includes a lysine residue known [2] to be essential for the enzyme's activity.

Consensus pattern: Q-W-G-D-E-G-K-G

Consensus pattern: G-I-[GR]-P-x-Y-x(2)-K-x(2)-R [K is the active site residue]

[1] Wiesmueller L., Wittbrodt J., Noegel A.A., Schleicher M. J. Biol. Chem. 266:2480-2485(1991).

[2] Silva M.M., Poland B.W., Hoffman C.R., Fromm H.J., Honzatko R.B. J. Mol. Biol. 254:431-446(1995).

[3] Bouyoub A., Barbier G., Forterre P., Labedan B. 2.3.CO;2-"J. Mol. Biol. 261:144-154(1996).

24. (AdoHcyase)

S-adenosyl-L-homocysteine hydrolase signatures

S-adenosyl-L-homocysteine hydrolase (EC 3.3.1.1) (AdoHcyase) is an enzyme of the activated methyl cycle, responsible for the reversible hydration of S-adenosyl-L-homocysteine into adenosine and homocysteine. AdoHcyase is a ubiquitous enzyme which binds and requires NAD⁺ as a cofactor. AdoHcyase is a highly conserved protein [1] of about

430 to 470 amino acids. Two highly conserved regions were selected as signature patterns. The first pattern is located in the N-terminal section; the second is derived from aglycine-rich region in the central part of AdoHcyase; a region thought to be involved in NAD-binding.

5 Consensus pattern: [GSA]-[CS]-N-x-[FYLM]-S-[ST]-[QA]-[DEN]-x-[AV]-[AT]-[AD]-[AC]-[LIVMCG]

Consensus pattern: [GA]-[KS]-x(3)-[LIV]-x-G-[FY]-G-x-[VC]-G-[KRL]-G-x-[ASC]

10 [1] Sganga M.W., Aksamit R.R., Cantoni G.L., Bauer C.E. Proc. Natl. Acad. Sci. U.S.A. 89:6328-6332(1992).

25. AhpC/TSA family

15 This family contains proteins related to alkyl hydroperoxide reductaseComment: (AhpC) and thiol specific antioxidant (TSA).

[1] Chae HZ, Robison K, Poole LB, Church G, Storz G, Rhee SG, Proc Natl Acad Sci U S A 1994;91:7017-7021

26. (Aldose epim)

25 Aldose 1-epimerase putative active site Aldose 1-epimerase (EC 5.1.3.3) (mutarotase) is the enzyme responsible for the anomeric interconversion of D-glucose and other aldoses between their alpha- and beta-forms. The sequence of mutarotase from two bacteria, *Acinetobacter calcoaceticus* and *Streptococcus thermophilus* is available [1]. It has also been shown that, on the basis of extensive sequence similarities, a mutarotase domain seem to be present in the C-terminal half of the fungal GAL10 protein which encodes, in the N-terminal

30 part, for UDP-glucose 4-epimerase. The best conserved region in the sequence of mutarotase is centered around a conserved histidine residue which may be involved in the catalytic mechanism.

Consensus pattern: [NS]-x-T-N-H-x-Y-[FW]-N-[LI]

[1] Poolman B., Royer T.J., Mainzer S.E., Schmidt B.F. J. Bacteriol. 172:4037-4047(1990).

5

27. (AlkA DNA repair)

Alkylbase DNA glycosidases alkA family signature

10

Alkylbase DNA glycosidases [1] are DNA repair enzymes that hydrolyzes the deoxyribose N-glycosidic bond to excise various alkylated bases from a damaged DNA polymer. In *Escherichia coli* there are two alkylbase DNA glycosidases: one (gene tag) which is constitutively expressed and which is specific for the removal of 3-methyladenine (EC 3.2.2.20), and one (gene alkA) which is induced during adaptation to alkylation and which can remove a variety of alkylation products (EC 3.2.2.21). Tag and alkA do not share any region of sequence similarity. In yeast there is an alkylbase DNA glycosidase (gene MAG1) [2,3], which can remove 3-methyladenine or 7-methyladenine and which is structurally related to alkA. MAG and alkA are both proteins of about 300 amino acid residues. While the C- and N-terminal ends appear to be unrelated, there is a central region of about 130 residues which is well conserved. A portion of this region has been selected as a signature pattern .

20

Consensus pattern: G-I-G-x-W-[ST]-[AV]-x-[LIVMFY](2)-x-[LIVM]-x(8)-[MF]-x(2)-[ED]-D

25

[1] Lindahl T., Sedgwick B. Annu. Rev. Biochem. 57:133-157(1988).

[2] Berdal K.G., Bjoras M., Bjelland S., Seeberg E.C. EMBO J. 9:4563-4568(1990).

[3] Chen J., Derfler B., Samson L. EMBO J. 9:4569-4575(1990).

30

28. Ammonium transporters signature

A number of proteins involved in the transport of ammonium ions across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionarily related. These

proteins are: - Yeast ammonium transporters MEP1, MEP2 and MEP3. - Arabidopsis thaliana high affinity ammonium transporter (gene AMT1). - Corynebacterium glutamicum ammonium and methylammonium transport system. - Escherichia coli putative ammonium transporter amtB. - Bacillus subtilis nrgA. - Mycobacterium tuberculosis hypothetical protein MtCY338.09c. - Synechocystis strain PCC 6803 hypothetical proteins slI0108, slI0537 and slI1017. - Methanococcus jannaschii hypothetical proteins MJ0058 and MJ1343. - Caenorhabditis elegans hypothetical proteins C05E11.4, F49E11.3 and M195.3. As expected by their transport function, these proteins are highly hydrophobic and seem to contain from 10 to 12 transmembrane domains. The best conserved region seems to be located in the fifth (or sixth) transmembrane region and is used as a signature pattern.

Consensus pattern: D-[FYWS]-A-G-[GSC]-x(2)-[IV]-x(3)-[SAG](2)-x(2)-[SAG]-[LIVMF]-x(3)-[LIVMFYWA](2)-x-[GK]-x-R

- [1] Ninnemann O., Janniaux J.-C., Frommer W.B. EMBO J. 13:3464-3471(1994).
- [2] Siewe R.M., Weil B., Burkovski A., Eikmanns B.J., Eikmanns M., Kraemer R. J. Biol. Chem. 271:5398-5403(1996).
- [3] Saier M.H. Jr. Adv. Microbiol. Physiol. 40:81-136(1998).

29. (Arch_histone)
CBF/NF-Y subunits signatures

Diverse DNA binding proteins are known to bind the CCAAT box, a common cis-acting element found in the promoter and enhancer regions of a large number of genes in eukaryotes. Amongst these proteins is one known as the CCAAT-binding factor (CBF) or NF-Y [1]. CBF is a heteromeric transcription factor that consists of two different components both needed for DNA-binding. The HAP protein complex of yeast binds to the upstream activation site of cytochrome C iso-1 gene (CYC1) as well as other genes involved in mitochondrial electron transport and activates their expression. It also recognizes the sequence CCAAT and is structurally and evolutionary related to CBF. The first subunit of CBF, known as CBF-A or NF-YB in vertebrates, HAP3 in budding yeast and as php3 in fission yeast, is a protein of 116 to 210 amino-acid residues which contains a highly

conserved central domain of about 90 residues. This domain seems to be involved in DNA-binding; a signature pattern had been developed from its central part. The second subunit of CBF, known as CBF-B or NF-YA in vertebrates, HAP2 in budding yeast and php2 in fission yeast, is a protein of 265 to 350 amino-acid residues which contains a highly conserved region of about 60 residues. This region, called the 'essential core' [2], seems to consist of two subdomains: an N-terminal subunit-association domain and a C-terminal DNA recognition domain. A signature pattern has been developed from a section of the subunit-association domain.

Consensus pattern: C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C-

Consensus pattern: Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E-

[1] Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I., Benoist C., Mathis D.

Nucleic Acids Res. 20:1087-1091(1992).

[2] Olesen J.T., Fikes J.D., Guarente L. Mol. Cell. Biol. 11:611-619(1991).

30. Argininosuccinate synthase signatures

Argininosuccinate synthase (EC 6.3.4.5) (AS) is a urea cycle enzyme that catalyzes the penultimate step in arginine biosynthesis: the ATP-dependent ligation of citrulline to aspartate to form argininosuccinate, AMP and pyrophosphate [1,2]. In humans, a defect in the AS gene causes citrullinemia, a genetic disease characterized by severe vomiting spells and mental retardation. AS is a homotetrameric enzyme of chains of about 400 amino-acid residues. An arginine seems to be important for the enzyme's catalytic mechanism. The sequences of AS from various prokaryotes, archaeobacteria and eukaryotes show significant similarity. Two signature patterns have been selected for AS. The first is a highly conserved stretch of nine residues located in the N-terminal extremity of these enzymes, the second is derived from a conserved region which contains one of the conserved arginine residues.

Consensus pattern: [AS]-[FY]-S-G-G-[LV]-D-T-[ST]-

77

Consensus pattern: G-x-T-x-K-G-N-D-x(2)-R-F-

[1] van Vliet F., Crabeel M., Boyen A., Tricot C., Stalon V., Falmagne P., Nakamura Y.,
Baumberg S., Glansdorff N. Gene 95:99-104(1990).

5 [2] Morris C.J., Reeve J.N. J. Bacteriol. 170:3125-3130(1988).

31. Armadillo/beta-catenin-like repeats

10 Approx. 40 amino acid repeat. Tandem repeats form super-helix of helices that is proposed to
mediate interaction of beta-catenin with its ligands. CAUTION: This family does not contain
all known armadillo repeats.

[1] Huber AH, Nelson WJ, Weis WI, Cell 1997;90:871-882.

[2] Gumbiner BM, Curr Opin Cell Biol 1995;7:634-640.

15 [3] Cavallo R, Rubenstein D, Peifer M, Curr Opin Genet Dev 1997;7:459-466.

[4] Su LK, Vogelstein B, Kinzler KW, Science 1993;262:1734-1737.

[5] Masiarz FR, Munemitsu S, Polakis P Science 1993;262:1731-1734

[6] Peifer M, Wieschaus E, Cell 1990;63:1167-1176.

32. (Asn Synthase)

Asparagine synthase

25 This family is always found associated with GATase_2. Members of this family catalyse the
conversion of aspartate to asparagine.

33. Asparaginase_2

Asparaginase 12 members

30

34. (Aspartyl tRNA N)

Aminoacyl-transfer RNA synthetases class-II signatures

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).

[2] Delarue M., Moras D. BioEssays 15:675-687(1993).

[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).

[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

[5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).

[6] Cusack S. Biochimie 75:1077-1081(1993).

[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).

[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

35. (ArfGap) Putative GTP-ase activating protein for Arf. Putative zinc fingers with GTPase activating proteins (GAPs) towards the small GTPase, Arf. The GAP of ARD1 stimulates

GTPase hydrolysis for ARD1 but not ARFs. Number of members: 34

[1]Medline: 96324970. Identification and cloning of centaurin- α . A novel phosphatidylinositol 3,4,5-trisphosphate-binding protein from rat brain. Hammonds-Odie LP, Jackson TR, Profit AA, Blader U, Turck CW, Prestwich GD, Theibert AB; J Biol Chem 1996;271:18859-18868.

[2]Medline: 97296423. A target of phosphatidylinositol 3,4,5-trisphosphate with a zinc finger motif similar to that of the ADP-ribosylation -factor GTPase-activating protein and two pleckstrin homology domains. Tanaka K, Imajoh-Ohmi S, Sawada T, Shirai R, Hashimoto Y, Iwasaki S, Kaibuchi K, Kanaho Y, Shirai T, Terada Y, Kimura K, Nagata S, Fukui Y; Eur J Biochem 1997;245:512-519.

[3] 98112795. Molecular characterization of the GTPase-activating domain of ADP-ribosylation factor domain protein 1 (ARD1). Vitale N, Moss J, Vaughan M; J Biol Chem 1998;273:2553-2560.

36. Apolipoprotein. Apolipoprotein A1/A4/E family. This family includes: Swiss:P02647 Apolipoprotein A-I. Swiss:P06727 Apolipoprotein A-IV. Swiss:P02649 Apolipoprotein E. These proteins contain several 22 residue repeats which form a pair of alpha helices. Number of members: 42

[1]Medline: 91289138. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA; Science 1991;252:1817-1822.

37. Amino acid permeases signature

Amino acid permeases are integral membrane proteins involved in the transport of amino acids into the cell. A number of such proteins have been found to be evolutionary related [1,2,3]. These proteins are: - Yeast general amino acid permeases (genes GAP1, AGP2 and AGP3). - Yeast basic amino acid permease (gene ALP1). - Yeast Leu/Val/Ile permease (gene BAP2). - Yeast arginine permease (gene CAN1). - Yeast dicarboxylic amino acid permease (gene DIP5). - Yeast asparagine/glutamine permease (gene AGP1). - Yeast glutamine

permease (gene GNP1). - Yeast histidine permease (gene HIP1). - Yeast lysine permease (gene LYP1). - Yeast proline permease (gene PUT4). - Yeast valine and tyrosine permease (gene VAL1/TAT1). - Yeast tryptophan permease (gene TAT2/SCM2). - Yeast choline transport protein (gene HNM1/CTR1). - Yeast GABA permease (gene UGA4). - Yeast hypothetical protein YKL174c. - Fission yeast protein isp5. - Fission yeast hypothetical protein SpAC8A4.11 - Fission yeast hypothetical protein SpAC11D3.08c. - *Emmericella nidulans* proline transport protein (gene prnB). - *Trichoderma harzianum* amino acid permease INDA1. - *Salmonella typhimurium* L-asparagine permease (gene ansP). - *Escherichia coli* aromatic amino acid transport protein (gene aroP). - *Escherichia coli* D-serine/D-alanine/glycine transporter (gene cycA). - *Escherichia coli* GABA permease (gene gabP). - *Escherichia coli* lysine-specific permease (gene lysP). - *Escherichia coli* phenylalanine-specific permease (gene pheP). - *Salmonella typhimurium* proline-specific permease (gene proY). - *Escherichia coli* and *Klebsiella pneumoniae* hypothetical protein yeeF. - *Escherichia coli* and *Salmonella typhimurium* hypothetical protein yifK. - *Bacillus subtilis* permeases rocC and rocE which probably transports arginine or ornithine. These proteins seem to contain up to 12 transmembrane segments. As a signature for this family of proteins, the best conserved region which is located in the second transmembrane segment has been selected.

Consensus pattern: [STAGC]-G-[PAG]-x(2,3)-[LIVMFYWA](2)-x-[LIVMFYW]-x-[LIVMFWSTAGC](2)-[STAGC]-x(3)-[LIVMFYWT]-x-[LIVMST]-x(3)-[LIVMCTA]-[GA]-E-x(5)-[PSAL]-

[1] Weber E., Chevalier M.R., Jund R. J. Mol. Evol. 27:341-350(1988).

[2] Vandenbol M., Jauniaux J.-C., Grenson M. Gene 83:153-159(1989).

[3] Reizer J., Finley K., Kakuda D., McLeod C.L., Reizer A., Saier M.H. Jr. Protein Sci. 2:20-30(1993).

38. aakinas (1) Glutamate 5-kinase signature

Glutamate 5-kinase (EC 2.7.2.11) (gamma-glutamyl kinase) (GK) is the enzyme that catalyzes the first step in the biosynthesis of proline from glutamate, the ATP-dependent phosphorylation of L-glutamate into L-glutamate 5-phosphate. In eubacteria (gene proB) and

yeast [1] (gene PRO1), GK is a monofunctional protein, while in plants and mammals, it is a bifunctional enzyme (P5CS) [2] that consists of two domains: a N-terminal GK domain and a C-terminal gamma-glutamyl phosphate reductase domain (EC 1.2.1.41) (see <PDOC00940>). As a signature pattern, a highly conserved glycine-and alanine-rich region located in the central section of these enzymes has been selected. Yeast hypothetical protein YHR033w is highly similar to GK.

Consensus pattern: [GSTN]-x(2)-G-x-G-[GC]-[IM]-x-[STA]-K-[LIVM]-x-[SA]-[TCA]-x(2)-[GALV]-x(3)-G-

[1] Li W., Brandriss M.C. J. Bacteriol. 174:4148-4156(1992).

[2] Hu C.-A.A., Delauney A.J., Verma D.P.S. Proc. Natl. Acad. Sci. U.S.A. 89:9354-9358(1992).

aakinase (2) Aspartokinase signature

Aspartokinase (EC 2.7.2.4) (AK) [1] catalyzes the phosphorylation of aspartate. The product of this reaction can then be used in the biosynthesis of lysine or in the pathway leading to homoserine, which participates in the biosynthesis of threonine, isoleucine and methionine. In *Escherichia coli*, there are three different isozymes which differ in their sensitivity to repression and inhibition by Lys, Met and Thr. AK1 (gene thrA) and AK2 (gene metL) are bifunctional enzymes which both consist of an N-terminal AK domain and a C-terminal homoserine dehydrogenase domain. AK1 is involved in threonine biosynthesis and AK2, in that of methionine. The third isozyme, AK3 (gene lysC), is monofunctional and involved in lysine synthesis. In yeast, there is a single isozyme of AK (gene HOM3). As a signature pattern for AK, a conserved region located in the N-terminal extremity has been selected.

Consensus pattern: [LIVM]-x-K-[FY]-G-G-[ST]-[SC]-[LIVM]-

[1] Rafalski J.A., Falco S.C. J. Biol. Chem. 263:2146-2151(1988).

aakinase (3) Gamma-glutamyl phosphate reductase signature

Gamma-glutamyl phosphate reductase (EC 1.2.1.41) (GPR) is the enzyme that catalyzes the second step in the biosynthesis of proline from glutamate, the NADP-dependent reduction of

L-glutamate 5-phosphate into L-glutamate 5-semialdehyde and phosphate. In eubacteria (gene proA) and yeast [1] (gene PRO2), GPR is a monofunctional protein, while in plants and mammals, it is a bifunctional enzyme (P5CS) [2] that consists of two domains: a N-terminal glutamate 5-kinase domain (EC 2.7.2.11) (see <PDOC00701>) and a C-terminal GPR domain. As a signature pattern, a conserved region that contains two histidine residues has been selected. This region is located in the last third of GPR.

Consensus pattern: V-x(5)-A-[LIV]-x-H-I-x(2)-[HY]-[GS]-[ST]-x-H-[ST]-[DE]-x-I-

[1] Pearson B.M., Hernando Y., Payne J., Wolf S.S., Kalogeropoulos A., Schweizer M. Yeast 12:1021-1031(1996).

[2] Hu C.-A.A., Delauney A.J., Verma D.P.S. Proc. Natl. Acad. Sci. U.S.A. 89:9354-9358(1992).

39. (abhydrolase) alpha/beta hydrolase fold. This catalytic domain is found in a very wide range of enzymes.

[1] Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolof F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschuere KHG, Goldman A, Protein Eng 1992;5:197-211.

40. (Acid phosphatase) Histidine acid phosphatases signatures

Acid phosphatases (EC 3.1.3.2) are a heterogeneous group of proteins that hydrolyze phosphate esters, optimally at low pH. It has been shown [1] that a number of acid phosphatases, from both prokaryotes and eukaryotes, share two regions of sequence similarity, each centered around a conserved histidine residue. These two histidines seem to be involved in the enzymes' catalytic mechanism [2,3]. The first histidine is located in the N-terminal section and forms a phosphohistidine intermediate while the second is located in the C-terminal section and possibly acts as proton donor. Enzymes belonging to this family are called 'histidine acid phosphatases' and are listed below:

- *Escherichia coli* glucose-1-phosphatase (EC 3.1.3.10) (gene *agp*).

- Yeast constitutive and repressible acid phosphatases (genes PHO3 and PHO5).

5 - Fission yeast acid phosphatase (gene *pho1*).

- *Aspergillus* phytases A and B (EC 3.1.3.8) (gene *phvA* and *phvB*).

- Mammalian lysosomal acid phosphatase.

- Mammalian prostatic acid phosphatase.

10 - *Caenorhabditis elegans* hypothetical proteins B0361.7, C05C10.1, C05C10.4
and F26C11.1.

Consensus pattern[LIVM]-x(2)-[LIVMA]-x(2)-[LIVM]-x-R-H-[GN]-x-R-x-[PAS] [H is the phosphohistidine residue]

11

15 Consensus pattern[LIVMF]-x-[LIVMFAG]-x(2)-[STAGI]-H-D-[STANQ]-x-[LIVM]-x(2)-
[LIVMFY]-x(2)-[STA] [H is an active site residue] Sequences known to belong to this class
detected by the patternALL, except for rat prostatic acid phosphatase which seems to have
Tyr instead of the active site His

20 [1] van Etten R.L., Davidson R., Stevis P.E., MacArthur H., Moore D.L. J. Biol. Chem. 266:2313-2319(1991).

[2] Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M., van Etten R.L. J. Biol. Chem. 267:22830-22836(1992).

[3] Schneider G., Lindqvist Y., Vihko P. EMBO J. 12:2609-2615(1993).

25

41. Aconitase family signatures

Aconitase (aconitate hydratase) (EC 4.2.1.3) [1] is the enzyme from the tricarboxylic acid cycle that catalyzes the reversible isomerization of citrate and isocitrate. Cis-aconitate is formed as an intermediary product during the course of the reaction. In eukaryotes two isozymes of aconitase are known to exist: one found in the mitochondrial matrix and the other found in the cytoplasm. Aconitase, in its active form, contains a 4Fe-4S iron-sulfur cluster; three cysteine residues have been shown to be ligands of the 4Fe-4S cluster. It has

been shown that the aconitase family also contains the following proteins: - Iron-responsive element binding protein (IRE-BP). IRE-BP is a cytosolic protein that binds to iron-responsive elements (IREs). IREs are stem-loop structures found in the 5'UTR of ferritin, and delta aminolevulinic acid synthase mRNAs, and in the 3'UTR of transferrin receptor mRNA. IRE-

5 BP also express aconitase activity. - 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate isomerase), the enzyme that catalyzes the second step in the biosynthesis of leucine. - Homoaconitase (EC 4.2.1.36) (homoaconitate hydratase), an enzyme that participates in the alpha-aminoadipate pathway of lysine biosynthesis and that converts cis-homoaconitate into homoisocitric acid. - *Escherichia coli* protein ybhJ

10 Consensus pattern: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-[LIVMA] [C binds the iron-sulfur center]

Consensus pattern: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA] [The two C's bind the iron-sulfur center]-

15 [1] Gruer M.J., Artymiuk P.J., Guest J.R. Trends Biochem. Sci. 22:3-6(1997).

42. Actins signatures

20 Actins [1 to 4] are highly conserved contractile proteins that are present in all eukaryotic cells. In vertebrates there are three groups of actin isoforms: alpha, beta and gamma. The alpha actins are found in muscle tissues and are a major constituent of the contractile apparatus. The beta and gamma actins co-exists in most cell types as components of the cytoskeleton and as mediators of internal cell motility. In plants [5] there are many isoforms

25 which are probably involved in a variety of functions such as cytoplasmic streaming, cell shape determination, tip growth, graviperception, cell wall deposition, etc. Actin exists either in a monomeric form (G-actin) or in a polymerized form (F-actin). Each actin monomer can bind a molecule of ATP; when polymerization occurs, the ATP is hydrolyzed. Actin is a protein of from 374 to 379 amino acid residues. The structure of actin has been highly

30 conserved in the course of evolution. Recently some divergent actin-like proteins have been identified in several species. These proteins are: - Contractin (actin-RPV) from mammals, fungi (yeast ACT5, *Neurospora crassa* ro-4) and *Pneumocystis carinii* (actin-II). Contractin seems to be a component of a multi-subunit centrosomal complex involved in microtubule

based vesicle motility. This subfamily is also known as ARP1. - ARP2 subfamily which includes chicken ACTL, yeast ACT2, Drosophila 14D, C.elegans actC. - ARP3 subfamily which includes actin 2 from mammals, Drosophila 66B, yeast ACT4 and fission yeast act2. - ARP4 subfamily which includes yeast ACT3 and Drosophila 13E. Three signature patterns have been developed. The first two are specific to actins and span positions 54 to 64 and 357 to 365. The last signature picks up both actins and the actin-like proteins and corresponds to positions 106 to 118 in actins.

Consensus pattern: [FY]-[LIV]-G-[DE]-E-A-Q-x-[RKQ](2)-G-

Consensus pattern: W-[IV]-[STA]-[RK]-x-[DE]-Y-[DNE]-[DE]-

Consensus pattern: [LM]-[LIVM]-T-E-[GAPQ]-x-[LIVMFYWHQ]-N-[PSTAQ]-x(2)-N-[KR]-

[1] Shterline P., Clayton J., Sparrow J.C. (In) Actins, 3rd Edition, Academic Press Ltd, London, (1996).

[2] Pollard T.D., Cooper J.A. Annu. Rev. Biochem. 55:987-1036(1986).

[3] Pollard T.D. Curr. Opin. Cell Biol. 1:33-40(1990).

[4] Rubenstein P.A. BioEssays 12:309-315(1990).

[5] Meagher R.B., McLean B.G. Cell Motil. Cytoskeleton 16:164-166(1990).

43. Adenylate kinase signature

Adenylate kinase (EC 2.7.4.3) (AK) [1] is a small monomeric enzyme that catalyzes the reversible transfer of MgATP to AMP ($\text{MgATP} + \text{AMP} = \text{MgADP} + \text{ADP}$). In mammals there are three different isozymes: - AK1 (or myokinase), which is cytosolic. - AK2, which is located in the outer compartment of mitochondria. - AK3 (or GTP:AMP phosphotransferase), which is located in the mitochondrial matrix and which uses MgGTP instead of MgATP. The sequence of AK has also been obtained from different bacterial species and from plants and fungi. Two other enzymes have been found to be evolutionary related to AK. These are: - Yeast uridylylate kinase (EC 2.7.4.-) (UK) (gene URA6) [2] which catalyzes the transfer of a phosphate group from ATP to UMP to form UDP and ADP. - Slime mold UMP-CMP kinase (EC 2.7.4.14) [3] which catalyzes the transfer of a phosphate group from ATP to either CMP or UMP to form CDP or UDP and ADP. Several regions of AK family enzymes are well

conserved, including the ATP-binding domains. The most conserved of all regions have been selected as a signature for this type of enzyme. This region includes an aspartic acid residue that is part of the catalytic cleft of the enzyme and that is involved in a salt bridge. It also includes an arginine residue whose modification leads to inactivation of the enzyme

Consensus pattern: [LIVMFYW](3)-D-G-[FYI]-P-R-x(3)-[NQ]-

[1] Schulz G.E. Cold Spring Harbor Symp. Quant. Biol. 52:429-439(1987).

[2] Liljelund P., Sanni A., Friesen J.D., Lacroute F. Biochem. Biophys. Res. Commun. 165:464-473(1989).

[3] Wiesmueller L., Noegel A.A., Barzu O., Gerisch G., Schleicher M. J. Biol. Chem. 265:6339-6345(1990).

[4] Kath T.H., Schmid R., Schaefer G. Arch. Biochem. Biophys. 307:405-410(1993).

44. (adh_short) Short-chain dehydrogenases/reductases family signature. The short-chain dehydrogenases/reductases family (SDR) [1] is a very large family of enzymes, most of which are known to be NAD- or NADP-dependent oxidoreductases. As the first member of this family to be characterized was *Drosophila* alcohol dehydrogenase, this family used to be called [2,3,4]'insect-type', or 'short-chain' alcohol dehydrogenases. Most member of this family are proteins of about 250 to 300 amino acid residues. The proteins currently known to belong to this family are listed below. - Alcohol dehydrogenase (EC 1.1.1.1) from insects such as *Drosophila*. - Acetoin dehydrogenase (EC 1.1.1.5) from *Klebsiella terrigena* (gene budC). - D-beta-hydroxybutyrate dehydrogenase (BDH) (EC 1.1.1.30) from mammals. - Acetoacetyl-CoA reductase (EC 1.1.1.36) from various bacterial species (gene phbB or phaB). - Glucose 1-dehydrogenase (EC 1.1.1.47) from *Bacillus*. - 3-beta-hydroxysteroid dehydrogenase (EC 1.1.1.51) from *Comomonas testosteroni*. - 20-beta-hydroxysteroid dehydrogenase (EC 1.1.1.53) from *Streptomyces hydrogenans*. - Ribitol dehydrogenase (EC 1.1.1.56) (RDH) from *Klebsiella aerogenes*. - Estradiol 17-beta-dehydrogenase (EC 1.1.1.62) from human. - Gluconate 5-dehydrogenase (EC 1.1.1.69) from *Gluconobacter oxydans* (gene gno). - 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) from *Escherichia coli* (gene fabG) and from plants. - Retinol dehydrogenase (EC 1.1.1.105) from mammals. - 2-deoxy-d-gluconate 3-dehydrogenase (EC 1.1.1.125) from *Escherichia coli* and *Erwinia chrysanthemi*

- (gene kduD). - Sorbitol-6-phosphate 2-dehydrogenase (EC [1.1.1.140](#)) from *Escherichia coli* (gene gutD) and from *Klebsiella pneumoniae* (gene sorD). - 15-hydroxyprostaglandin dehydrogenase (NAD⁺) (EC [1.1.1.141](#)) from human. - Corticosteroid 11-beta-dehydrogenase (EC [1.1.1.146](#)) (11-DH) from mammals. - 7-alpha-hydroxysteroid dehydrogenase (EC [1.1.1.159](#)) from *Escherichia coli* (gene hdhA), *Eubacterium* strain VPI 12708 (gene baiA) and from *Clostridium sordellii*. - NADPH-dependent carbonyl reductase (EC [1.1.1.184](#)) from mammals. - Tropinone reductase-I (EC [1.1.1.206](#)) and -II (EC [1.1.1.236](#)) from plants. - N-acylmannosamine 1-dehydrogenase (EC [1.1.1.233](#)) from *Flavobacterium* strain 141-8. - D-arabinitol 2-dehydrogenase (ribulose forming) (EC [1.1.1.250](#)) from fungi. - Tetrahydroxynaphthalene reductase (EC [1.1.1.252](#)) from *Magnaporthe grisea*. - Pteridine reductase 1 (EC [1.1.1.253](#)) (gene PTR1) from *Leishmania*. - 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-) from *Pseudomonas paucimobilis*. - Cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase (EC 1.3.1.-) from *Acinetobacter calcoaceticus* (gene benD) and *Pseudomonas putida* (gene xylL). - Biphenyl-2,3-dihydro-2,3-diol dehydrogenase (EC 1.3.1.-) (gene bphB) from various *Pseudomonaceae*. - Cis-toluene dihydrodiol dehydrogenase (EC 1.3.1.-) from *Pseudomonas putida* (gene todD). - Cis-benzene glycol dehydrogenase (EC [1.3.1.19](#)) from *Pseudomonas putida* (gene bnzE). - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC [1.3.1.28](#)) from *Escherichia coli* (gene entA) and *Bacillus subtilis* (gene dhbA). - Dihydropteridine reductase (EC [1.6.99.7](#)) (HDHPR) from mammals. - Lignin degradation enzyme ligD from *Pseudomonas paucimobilis*. - Agropine synthesis reductase from *Agrobacterium* plasmids (gene masI). - Versicolorin reductase from *Aspergillus parasiticus* (gene VER1). - Putative keto-acyl reductases from *Streptomyces* polyketide biosynthesis operons. - A trifunctional hydratase-dehydrogenase-epimerase from the peroxisomal beta-oxidation system of *Candida tropicalis*.
- This protein contains two tandemly repeated 'short-chain dehydrogenase-type' domain in its N-terminal extremity. - Nodulation protein nodG from species of *Azospirillum* and *Rhizobium* which is probably involved in the modification of the nodulation Nod factor fatty acyl chain. - Nitrogen fixation protein fixR from *Bradyrhizobium japonicum*. - *Bacillus subtilis* protein dltE which is involved in the biosynthesis of D- alanyl-lipoteichoic acid. - Human follicular variant translocation protein 1 (FVT1). - Mouse adipocyte protein p27. - Mouse protein Kc 6. - Maize sex determination protein TASSELSEED 2. - *Sarcophaga peregrina* 25 Kd development specific protein. - *Drosophila* fat body protein P6. - A *Listeria monocytogenes* hypothetical protein encoded in the internalins gene region. - *Escherichia coli*

hypothetical protein yciK. - Escherichia coli hypothetical protein ydfG. - Escherichia coli
 hypothetical protein yjgI. - Escherichia coli hypothetical protein yjgU. - Escherichia coli
 hypothetical protein yohF. - Bacillus subtilis hypothetical protein yoxD. - Bacillus subtilis
 hypothetical protein ywF. - Bacillus subtilis hypothetical protein ywH. - Yeast hypothetical
 5 protein YIL124w. - Yeast hypothetical protein YIR035c. - Yeast hypothetical protein
 YIR036c. - Yeast hypothetical protein YKL055c. - Fission yeast hypothetical protein
 SpAC23D3.11. One of the best conserved regions which includes two perfectly conserved
 residues, a tyrosine and a lysine has been selected as a signature pattern for this family of
 proteins. The tyrosine residue participates in the catalytic mechanism.

10 Consensus pattern: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCI VM]-[STAGC]-K-
 {PC}-[SAGFYR]-[LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)-[LIVMFYWGAPTHQ]-
 [GSACQRHM] [Y is an active site residue] -

15 [1] Joernvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D.
 Biochemistry 34:6003-6013(1995).
 [2] Villarroya A., Juan E., Egestad B., Joernvall H. Eur. J. Biochem. 180:191-197(1989).
 [3] Persson B., Krook M., Joernvall H. Eur. J. Biochem. 200:537-543(1991).
 [4] Neidle E.L., Hartnett C., Ornston N.L., Bairoch A., Reik M., Harayama S. Eur. J.
 20 Biochem. 204:113-120(1992).

45. (adh_short_C2) Short-chain dehydrogenases/reductases family signature

The short-chain dehydrogenases/reductases family (SDR) [1] is a very large family of
 25 enzymes, most of which are known to be NAD- or NADP-dependent oxidoreductases. As the
 first member of this family to be characterized was Drosophila alcohol dehydrogenase, this
 family used to be called [2,3,4] 'insect-type', or 'short-chain' alcohol dehydrogenases. Most
 member of this family are proteins of about 250 to 300 amino acid residues. The proteins
 currently known to belong to this family are listed below. - Alcohol dehydrogenase (EC
 30 1.1.1.1) from insects such as Drosophila. - Acetoin dehydrogenase (EC 1.1.1.5) from
 Klebsiella terrigena (gene budC). - D-beta-hydroxybutyrate dehydrogenase (BDH) (EC
1.1.1.30) from mammals. - Acetoacetyl-CoA reductase (EC 1.1.1.36) from various bacterial
 species (gene phbB or phaB). - Glucose 1-dehydrogenase (EC 1.1.1.47) from Bacillus. - 3-

protein nodG from species of *Azospirillum* and *Rhizobium* which is probably involved in the modification of the nodulation Nod factor fatty acyl chain. - Nitrogen fixation protein fixR from *Bradyrhizobium japonicum*. - *Bacillus subtilis* protein dltE which is involved in the biosynthesis of D- alanyl-lipoteichoic acid. - Human follicular variant translocation protein 1 (FVT1). - Mouse adipocyte protein p27. - Mouse protein Ke 6. - Maize sex determination protein TASSELSEED 2. - *Sarcophaga peregrina* 25 Kd development specific protein. - *Drosophila* fat body protein P6. - A *Listeria monocytogenes* hypothetical protein encoded in the internalins gene region. - *Escherichia coli* hypothetical protein yciK. - *Escherichia coli* hypothetical protein ydfG. - *Escherichia coli* hypothetical protein yjgI. - *Escherichia coli* hypothetical protein yjgU. - *Escherichia coli* hypothetical protein yohF. - *Bacillus subtilis* hypothetical protein yoxD. - *Bacillus subtilis* hypothetical protein ywfD. - *Bacillus subtilis* hypothetical protein ywfH. - Yeast hypothetical protein YIL124w. - Yeast hypothetical protein YIR035c. - Yeast hypothetical protein YIR036c. - Yeast hypothetical protein YKL055c. - Fission yeast hypothetical protein SpAC23D3.11. One of the best conserved regions which includes two perfectly conserved residues, a tyrosine and a lysine has been used as a signature pattern for this family of proteins. The tyrosine residue participates in the catalytic mechanism.

Consensus pattern: [LIVSPADNK]-x(12)-Y-[PSTAGNVCV]-[STAGNQCI VM]-[STAGC]-K-{PC}-[SAGFYR]-[LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)-[LIVMFYWGAPTHQ]-[GSACORHM] [Y is an active site residue]

[1] Joernvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D. *Biochemistry* 34:6003-6013(1995).

[2] Villarroya A., Juan E., Egestad B., Joernvall H. *Eur. J. Biochem.* 180:191-197(1989).

[3] Persson B., Krook M., Joernvall H. *Eur. J. Biochem.* 200:537-543(1991).

[4] Needle E.L., Hartnett C., Ornston N.L., Bairoch A., Reikik M., Harayama S. *Eur. J. Biochem.* 204:113-120(1992).

46. (adh_zinc) Zinc-containing alcohol dehydrogenases signatures

Alcohol dehydrogenase (EC 1.1.1.1) (ADH) catalyzes the reversible oxidation of ethanol to acetaldehyde with the concomitant reduction of NAD [1]. Currently three, structurally and

catalytically, different types of alcohol dehydrogenases are known: - Zinc-containing 'long-chain' alcohol dehydrogenases. - Insect-type, or 'short-chain' alcohol dehydrogenases. - Iron-containing alcohol dehydrogenases. Zinc-containing ADH's [2,3] are dimeric or tetrameric enzymes that bind two atoms of zinc per subunit. One of the zinc atom is essential for catalytic activity while the other is not. Both zinc atoms are coordinated by either cysteine or histidine residues; the catalytic zinc is coordinated by two cysteines and one histidine. Zinc-containing ADH's are found in bacteria, mammals, plants, and in fungi. In most species there are more than one isozyme (for example, human have at least six isozymes, yeast have three, etc.). A number of other zinc-dependent dehydrogenases are closely related to zinc ADH [4], these are: - Xylitol dehydrogenase (EC 1.1.1.9) (D-xylulose reductase). - Sorbitol dehydrogenase (EC 1.1.1.14). - Aryl-alcohol dehydrogenase (EC 1.1.1.90) (benzyl alcohol dehydrogenase). - Threonine 3-dehydrogenase (EC 1.1.1.103). - Cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) (CAD) [5]. CAD is a plant enzyme involved in the biosynthesis of lignin. - Galactitol-1-phosphate dehydrogenase (EC 1.1.1.251). - Pseudomonas putida 5-exo-alcohol dehydrogenase (EC 1.1.1.-) [6]. - Escherichia coli starvation sensing protein rspB. - Escherichia coli hypothetical protein yjgB. - Escherichia coli hypothetical protein yjgV. - Escherichia coli hypothetical protein yjjN. - Yeast hypothetical protein YAL060w (FUN49). - Yeast hypothetical protein YAL061w (FUN50). - Yeast hypothetical protein YCR105w. The pattern that has been developed to detect this class of enzymes is based on a conserved region that includes a histidine residue which is the second ligand of the catalytic zinc atom. This family also includes NADP-dependent quinone oxidoreductase (EC 1.6.5.5), an enzyme found in bacteria (gene qor), in yeast and in mammals where, in some species such as rodents, it has been recruited as an eye lens protein and is known as zeta-crystallin [7]. The sequence of quinone oxidoreductase is distantly related to that other zinc-containing alcohol dehydrogenases and it lacks the zinc-ligand residues. The torpedo fish and mammalian synaptic vesicle membrane protein vat-1 is related to qor. A specific pattern has been developed for this subfamily.

Consensus pattern: G-H-E-x(2)-G-x(5)-[GA]-x(2)-[IVSAC] [H is a zinc ligand]

Consensus pattern: [GSD]-[DEQH]-x(2)-L-x(3)-[SA](2)-G-G-x-G-x(4)-Q-x(2)-[KR]-

[1] Branden C.-I., Joernvall H., Eklund H., Furugren B. (In) The Enzymes (3rd edition) 11:104-190(1975).

- [2] Joernvall H., Persson B., Jeffery J. Eur. J. Biochem. 167:195-201(1987).
- [3] Sun H.-W., Plapp B.V. J. Mol. Evol. 34:522-535(1992).
- [4] Persson B., Hallborn J., Walfridsson M., Hahn-Haegerdal B., Keraenen S., Penttilae M., Joernvall H. FEBS Lett. 324:9-14(1993).
- 5 [5] Knight M.E., Halpin C., Schuch W. Plant Mol. Biol. 19:793-801(1992).
- [6] Koga H., Aramaki H., Yamaguchi E., Takeuchi K., Horiuchi T., Gunsalus I.C. J. Bacteriol. 166:1089-1095(1986).
- [7] Joernvall H., Persson B., Du Bois G., Lavers G.C., Chen J.H., Gonzalez P., Rao P.V., Zigler J.S. Jr. FEBS Lett. 322:240-244(1993).

47. (aldehyd) Aldehyde dehydrogenases active sites

Aldehyde dehydrogenases (EC [1.2.1.3](#) and EC [1.2.1.5](#)) are enzymes which oxidize a wide variety of aliphatic and aromatic aldehydes. In mammals at least four different forms of the enzyme are known [1]: class-1 (or Ald C) a tetrameric cytosolic enzyme, class-2 (or Ald M) a tetrameric mitochondrial enzyme, class-3 (or Ald D) a dimeric cytosolic enzyme, and class IV a microsomal enzyme. Aldehyde dehydrogenases have also been sequenced from fungal and bacterial species. A number of enzymes are known to be evolutionary related to aldehyde dehydrogenases; these enzymes are listed below. - Plants and bacterial betaine-aldehyde dehydrogenase (EC [1.2.1.8](#)) [2], an enzyme that catalyzes the last step in the biosynthesis of betaine. - Plants and bacterial NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC [1.2.1.9](#)). - *Escherichia coli* succinate-semialdehyde dehydrogenase (NADP+) (EC [1.2.1.16](#)) (gene *gabD*) [3], which reduces succinate semialdehyde into succinate. - *Escherichia coli* lactaldehyde dehydrogenase (EC [1.2.1.22](#)) (gene *ald*) [4]. - Mammalian succinate semialdehyde dehydrogenase (NAD+) (EC [1.2.1.24](#)). - *Escherichia coli* phenylacetaldehyde dehydrogenase (EC [1.2.1.39](#)). - *Escherichia coli* 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (gene *hpcC*). - *Pseudomonas putida* 2-hydroxymuconic semialdehyde dehydrogenase [5] (genes *dmpC* and *xylG*), an enzyme in the meta-cleavage pathway for the degradation of phenols, cresols and catechol. - Bacterial and mammalian methylmalonate-semialdehyde dehydrogenase (MMSDH) (EC [1.2.1.27](#)) [6], an enzyme involved in the distal pathway of valine catabolism. - Yeast delta-1-pyrroline-5-carboxylate dehydrogenase (EC [1.5.1.12](#)) [7] (gene *PUT2*), which converts proline to glutamate. - Bacterial multifunctional *putA* protein, which contains a delta-1-pyrroline- 5-

carboxylate dehydrogenase domain. - 26G, a garden pea protein of unknown function which is induced by dehydration of shoots [8]. - Mammalian formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) [9]. This is a cytosolic enzyme responsible for the NADP-dependent decarboxylative reduction of 10-formyltetrahydrofolate into tetrahydrofolate. It is an protein of about 900 amino acids which consist of three domains; the C- terminal domain (480 residues) is structurally and functionally related to aldehyde dehydrogenases. - Yeast hypothetical protein YBR006w. - Yeast hypothetical protein YER073w. - Yeast hypothetical protein YHR039c. - *Caenorhabditis elegans* hypothetical protein F01F1.6.A glutamic acid and a cysteine residue have been implicated in the catalytic activity of mammalian aldehyde dehydrogenase. These residues are conserved in all the enzymes of this family. Two patterns have been derived for this family, one for each of the active site residues.

Consensus pattern: [LIVMFGA]-E-[LIMSTAC]-[GS]-G-[KNLM]-[SADN]-[TAPFV] [E is the active site residue]-

Consensus pattern: [FYLVA]-x(3)-G-[QE]-x-C-[LIVMGSTANC]-[AGCN]-x-[GSTADNEKR] [C is the active site residue

- [1] Hempel J., Harper K., Lindahl R. *Biochemistry* 28:1160-1167(1989).
- [2] Weretilnyk E.A., Hanson A.D. *Proc. Natl. Acad. Sci. U.S.A.* 87:2745-2749(1990).
- [3] Niegemann E., Schulz A., Bartsch K. *Arch. Microbiol.* 160:454-460(1993).
- [4] Hidalgo E., Chen Y.-M., Lin E.C.C., Aguilar J. J. *Bacteriol.* 173:6118-6123(1991).
- [5] Nordlund I., Shingler V. *Biochim. Biophys. Acta* 1049:227-230(1990).
- [6] Steele M.L., Lorenz D., Hatter K., Park A., Sokatch J.R. *J. Biol. Chem.* 267:13585-13592(1992).
- [7] Krzywicki K.A., Brandriss M.C. *Mol. Cell. Biol.* 4:2837-2842(1984).
- [8] Guerrero F.D., Jones J.T., Mullet J.E. *Plant Mol. Biol.* 15:11-26(1990).
- [9] Cook R.J., Lloyd R.S., Wagner C. J. *Biol. Chem.* 266:4965-4973(1991).

48. Aldo/keto reductase family signatures

The aldo-keto reductase family [1,2] groups together a number of structurally and functionally related NADPH-dependent oxidoreductases as well as some other proteins. The proteins known to belong to this family are: - Aldehyde reductase (EC 1.1.1.2). - Aldose

reductase (EC 1.1.1.21). - 3-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.50), which terminates androgen action by converting 5-alpha-dihydrotestosterone to 3-alpha-androstenediol. - Prostaglandin F synthase (EC 1.1.1.188) which catalyzes the reduction of prostaglandins H2 and D2 to F2-alpha. - D-sorbitol-6-phosphate dehydrogenase (EC 1.1.1.200) from apple. - Morphine 6-dehydrogenase (EC 1.1.1.218) from *Pseudomonas putida* plasmid pMDH7.2 (gene morA). - Chlordecone reductase (EC 1.1.1.225) which reduces the pesticide chlordecone (kepone) to the corresponding alcohol. - 2,5-diketo-D-gluconic acid reductase (EC 1.1.1.-) which catalyzes the reduction of 2,5-diketogluconic acid to 2-keto-L-gulonic acid, a key intermediate in the production of ascorbic acid. - NAD(P)H-dependent xylose reductase (EC 1.1.1.-) from the yeast *Pichia stipitis*. This enzyme reduces xylose into xylitol. - Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase (EC 1.3.1.20). - 3-oxo-5-beta-steroid 4-dehydrogenase (EC 1.3.99.6) which catalyzes the reduction of delta(4)-3-oxosteroids. - A soybean reductase, which co-acts with chalcone synthase in the formation of 4,2',4'-trihydroxychalcone. - Frog eye lens rho crystallin. - Yeast GCY protein, whose function is not known. - *Leishmania* major P110/11E protein. P110/11E is a developmentally regulated protein whose abundance is markedly elevated in promastigotes compared with amastigotes. Its exact function is not yet known. - *Escherichia coli* hypothetical protein yafB. - *Escherichia coli* hypothetical protein yghE. - Yeast hypothetical protein YBR149w. - Yeast hypothetical protein YHR104w. - Yeast hypothetical protein YJR096w. These proteins have all about 300 amino acid residues. Three consensus patterns have been developed that are specific to this family of proteins. The first one is located in the N-terminal section of these proteins. The second pattern is located in the central section. The third pattern, located in the C-terminal, is centered on a lysine residue whose chemical modification, in aldose and aldehydereductases, affect the catalytic efficiency.

Consensus pattern: G-[FY]-R-[HSAL]-[LIVMF]-D-[STAGC]-[AS]-x(5)-E-x(2)-[LIVM]-G -
 Consensus pattern: [LIVMFY]-x(9)-[KREQ]-x-[LIVM]-G-[LIVM]-[SC]-N-[FY]-
 Consensus pattern: [LIVM]-[PAIV]-[KR]-[ST]-x(4)-R-x(2)-[GSTAEQK]-[NSL]-x(2)-
 [LIVMFA] [K is a putative active site residue]-

[1] Bohren K.M., Bullock B., Wermuth B., Gabbay K.H. J. Biol. Chem. 264:9547-9551(1989).

[2] Bruce N.C., Willey D.L., Coulson A.F.W., Jeffery J. Biochem. J. 299:805-811(1994).

49. Alpha amylase. This family is classified as family 13 of the glycosyl hydrolases. The structure is an 8 stranded alpha/beta barrel, interrupted by a ~70 a.a. calcium-binding domain protruding between beta strand 3 and alpha helix 3, and a carboxyl-terminal Greek key beta-barrel domain.

[1] Larson SB, Greenwood A, Cascio D, Day J, McPherson A, J Mol Biol 1994;235:1560-1584.

50. Aminotransferases class-I pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-I, currently consists of the following enzymes: - Aspartate aminotransferase (AAT) (EC 2.6.1.1). AAT catalyzes the reversible transfer of the amino group from L-aspartate to 2-oxoglutarate to form oxaloacetate and L-glutamate. In eukaryotes, there are two AAT isozymes: one is located in the mitochondrial matrix, the second is cytoplasmic. In prokaryotes, only one form of AAT is found (gene aspC). - Tyrosine aminotransferase (EC 2.6.1.5) which catalyzes the first step in tyrosine catabolism by reversibly transferring its amino group to 2- oxoglutarate to form 4-hydroxyphenylpyruvate and L-glutamate. - Aromatic aminotransferase (EC 2.6.1.57) involved in the synthesis of Phe, Tyr, Asp and Leu (gene tyrB). - 1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) (ACC synthase) from plants. ACC synthase catalyzes the first step in ethylene biosynthesis. - Pseudomonas denitrificans cobC, which is involved in cobalamin biosynthesis. - Yeast hypothetical protein YJL060w. The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [GS]-[LIVMFYTAC]-[GSTA]-K-x(2)-[GSALVN]-[LIVMFA]-x-[GNAR]- x-R-[LIVMA]-[GA] [K is the pyridoxal-P attachment site]

[1] Bairoch A. Unpublished observations (1992).

[2] Sung M.H., Tanizawa K., Tanaka H., Kuramitsu S., Kagamiyama H., Hirotsu K., Okamoto A., Higuchi T., Soda K. J. Biol. Chem. 266:2567-2572(1991).

5

51. Aminotransferases class-II pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1] into subfamilies. One of these, called class-II, currently consists of the following enzymes: - Glycine acetyltransferase (EC 2.3.1.29), which catalyzes the addition of acetyl-CoA to glycine to form 2-amino-3-oxobutanoate (gene kbl). - 5-aminolevulinic acid synthase (EC 2.3.1.37) (delta-ALA synthase), which catalyzes the first step in heme biosynthesis via the Shemin (or C4) pathway, i.e. the addition of succinyl-CoA to glycine to form 5-aminolevulinate. - 8-amino-7-oxononanoate synthase (EC 2.3.1.47) (7-KAP synthetase), a bacterial enzyme (gene bioF) which catalyzes an intermediate step in the biosynthesis of biotin: the addition of 6-carboxy-hexanoyl-CoA to alanine to form 8-amino-7-oxononanoate. - Histidinol-phosphate aminotransferase (EC 2.6.1.9), which catalyzes the eighth step in histidine biosynthetic pathway: the transfer of an amino group from 3-(imidazol-4-yl)-2-oxopropyl phosphate to glutamic acid to form histidinol phosphate and 2-oxoglutarate. - Serine palmitoyltransferase (EC 2.3.1.50) from yeast (genes LCB1 and LCB2), which catalyzes the condensation of palmitoyl-CoA and serine to form 3- ketosphinganine. The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern

25

Consensus pattern: T-[LIVMFYW]-[STAG]-K-[SAG]-[LIVMFYWR]-[SAG]-x(2)-[SAG]
[K is the pyridoxal-P attachment site]-

[1] Bairoch A. Unpublished observations (1991).

30

52. Aminotransferases class-III pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-III, currently consists of the following

5 enzymes: - Acetylornithine aminotransferase (EC 2.6.1.11) which catalyzes the transfer of an amino group from acetylornithine to alpha-ketoglutarate, yielding N-acetyl-glutamic-5-semi-aldehyde and glutamic acid. - Ornithine aminotransferase (EC 2.6.1.13), which catalyzes the transfer of an amino group from ornithine to alpha-ketoglutarate, yielding glutamic-5- semi-aldehyde and glutamic acid. - Omega-amino acid--pyruvate aminotransferase (EC 2.6.1.18),

10 which catalyzes transamination between a variety of omega-amino acids, mono- and diamines, and pyruvate. It plays a pivotal role in omega amino acids metabolism. - 4-aminobutyrate aminotransferase (EC 2.6.1.19) (GABA transaminase), which catalyzes the transfer of an amino group from GABA to alpha-ketoglutarate, yielding succinate semialdehyde and glutamic acid. - DAPA aminotransferase (EC 2.6.1.62), a bacterial enzyme

15 (gene bioA) which catalyzes an intermediate step in the biosynthesis of biotin, the transamination of 7-keto-8-aminopelargonic acid (7-KAP) to form 7,8- diaminopelargonic acid (DAPA). - 2,2-dialkylglycine decarboxylase (EC 4.1.1.64), a *Pseudomonas cepacia* enzyme (gene dgdA) that catalyzes the decarboxylating amino transfer of 2,2-dialkylglycine and pyruvate to dialkyl ketone, alanine and carbon dioxide. - Glutamate-1-semialdehyde

20 aminotransferase (EC 5.4.3.8) (GSA). GSA is the enzyme involved in the second step of porphyrin biosynthesis, via the C5 pathway. It transfers the amino group on carbon 2 of glutamate-1- semialdehyde to the neighbouring carbon, to give delta-aminolevulinic acid. - *Bacillus subtilis* aminotransferase yhxA. - *Bacillus subtilis* aminotransferase yodT. - *Haemophilus influenzae* aminotransferase HI0949. - *Caenorhabditis elegans* aminotransferase

25 T01B11.2. The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [LIVMFYWC](2)-x-D-E-[IVA]-x(2)-G-[LIVMFAGC]-x(0,1)-[RSACLI]-x-[GSAD]-x(12,16)-D-[LIVMFC]-[LIVMFYSTA]-x(2)- [GSA]-K-x(3)-[GSTADNV]-[GSAC] [K is the pyridoxal-P attachment site]-

30

[1] Bairoch A. Unpublished observations (1992).[2] Yonaha K., Nishie M., Aibara S. J. Biol. Chem. 267:12506-12510(1992).

53. Ank repeat. There's no clear separation between noise and signal on the HMM search
Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures.

5 The repeats associate to form a higher order structure.

[1] A, Holak TA, FEBS Lett 1997;401:127-132.

[2] Lux SE, John KM, Bennett V, Nature 1990;345:736-739.

10 54. Aminotransferases class-IV signature

Aminotransferases share certain mechanistic features with other pyridoxal-phosphate dependent enzymes, such as the covalent binding of the pyridoxal-phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped
15 [1,2] into subfamilies. One of these, called class-IV, currently consists of the following enzymes:

- Branched-chain amino-acid aminotransferase (EC 2.6.1.42) (transaminase B), a bacterial (gene *ilvE*) and eukaryotic enzyme which catalyzes the reversible transfer of an amino group from 4-methyl-2-oxopentanoate to glutamate, to form leucine and 2-oxoglutarate.
- D-alanine aminotransferase (EC 2.6.1.21). A bacterial enzyme which catalyzes the transfer of the amino group from D-alanine (and other D-amino acids) to 2-oxoglutarate, to form pyruvate and D-aspartate.
- 4-amino-4-deoxychorismate (ADC) lyase (gene *pabC*). A bacterial enzyme that
25 converts ADC into 4-aminobenzoate (PABA) and pyruvate.

The above enzymes are proteins of about 270 to 415 amino-acid residues that share a few regions of sequence similarity. Surprisingly, the best-conserved region does not include the lysine residue to which the pyridoxal-phosphate group is known to be attached, in *ilvE*.

The region that has been selected as a signature pattern is located some 40 residues at the C-terminus side of the PIP-lysine
30

Consensus pattern: E-x-[STAGCI]-x(2)-N-[LIVMFAC]-[FY]-x(6,12)-[LIVMF]-x-T- x(6,8)-[LIVM]-x-[GS]-[LIVM]-x-[KR]-

[1] Green J.M., Merkel W.K., Nichols B.P. J. Bacteriol. 174:5317-5323(1992).

[2] Bairoch A. Unpublished observations (1992).

55. Aminotransferases class-V pyridoxal-phosphate attachment site

Aminotransferases share certain mechanistic features with other pyridoxal- phosphate dependent enzymes, such as the covalent binding of the pyridoxal- phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped [1,2] into subfamilies. One of these, called class-V, currently consists of the following enzymes: - Phosphoserine aminotransferase (EC 2.6.1.52), an enzyme which catalyzes the reversible interconversion of phosphoserine and 2-oxoglutarate to 3-phosphonooxypyruvate and glutamate. It is required both in the major phosphorylated pathway of serine biosynthesis and in pyridoxine biosynthesis. The bacterial enzyme (gene serC) is highly similar to a rabbit endometrial progesterone-induced protein (EPIP), which is probably a phosphoserine aminotransferase [3]. - Serine--glyoxylate aminotransferase (EC 2.6.1.45) (SGAT) (gene sgaA) from *Methylobacterium extorquens*. - Serine--pyruvate aminotransferase (EC 2.6.1.51). This enzyme also acts as an alanine--glyoxylate aminotransferase (EC 2.6.1.44). In vertebrates, it is located in the peroxisomes and/or mitochondria. - Isopenicillin N epimerase (gene cefD). This enzyme is involved in the biosynthesis of cephalosporin antibiotics and catalyzes the reversible isomerization of isopenicillin N and penicillin N. - NifS, a protein of the nitrogen fixation operon of some bacteria and cyanobacteria. The exact function of nifS is not yet known. A highly similar protein has been found in fungi (gene NFS1 or SPL1). - The small subunit of cyanobacterial soluble hydrogenase (EC 1.12.-.-). - Hypothetical protein ycbU from *Bacillus subtilis*. - Hypothetical protein YFL030w from yeast. The sequence around the pyridoxal-phosphate attachment site of this class of enzyme is sufficiently conserved to allow the creation of a specific pattern.

Consensus pattern: [LIVFYCHT]-[DGH]-[LIVMFYAC]-[LIVMFYA]-x(2)-[GSTAC]-[GSTA]- [HQR]-K-x(4,6)-G-x-[GSAT]-x-[LIVMFYSAC] [K is the pyridoxal-P attachment site]-

[1] Ouzounis C., Sander C. FEBS Lett. 322:159-164(1993).

[2] Bairoch A. Unpublished observations (1992).

[3] van der Zel A., Lam H.-M., Winkler M.E. *Nucleic Acids Res.* 17:8379-8379(1989).

56. Annexins repeated domain signature

Annexins [1 to 6] are a group of calcium-binding proteins that associate reversibly with membranes. They bind to phospholipid bilayers in the presence of micromolar free calcium concentration. The binding is specific for calcium and for acidic phospholipids. Annexins have been claimed to be involved in cytoskeletal interactions, phospholipase inhibition, intracellular signalling, anticoagulation, and membrane fusion. Each of these proteins consist of an N-terminal domain of variable length followed by four or eight copies of a conserved segment of sixty one residues. The repeat (sometimes known as an 'endonexin fold') consists of five alpha-helices that are wound into a right-handed superhelix [7]. The proteins known to belong to the annexin family are listed below: - Annexin I (Lipocortin 1) (Calpactin 2) (p35) (Chromobindin 9). - Annexin II (Lipocortin 2) (Calpactin 1) (Protein I) (p36) (Chromobindin 8). - Annexin III (Lipocortin 3) (PAP-III). - Annexin IV (Lipocortin 4) (Endonexin 1) (Protein II) (Chromobindin 4). - Annexin V (Lipocortin 5) (Endonexin 2) (VAC-alpha) (Anchoring CII) (PAP-I). - Annexin VI (Lipocortin 6) (Protein III) (Chromobindin 20) (p68) (p70). This is the only known annexin that contains 8 (instead of 4) repeats. - Annexin VII (Synexin). - Annexin VIII (Vascular anticoagulant-beta) (VAC-beta). - Annexin IX from *Drosophila*. - Annexin X from *Drosophila*. - Annexin XI (Calcyclin-associated annexin) (CAP-50). - Annexin XII from *Hydra vulgaris*. - Annexin XIII (Intestine-specific annexin) (ISA). The signature pattern for this domain spans positions 9 to 61 of the repeat and includes the only perfectly conserved residue (an arginine in position 22)-

Consensus pattern: [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]-x(7)-[LIVMF]-x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF]-

[1] Raynal P., Pollard H.B. *Biochim. Biophys. Acta* 1197:63-93(1994).

[2] Barton G.J., Newman R.H., Freemont P.S., Crumpton M.J. *Eur. J. Biochem.* 198:749-760(1991).

[3] Burgoyne R.D., Geisow M.J. *Cell Calcium* 10:1-10(1989).

[4] Haigler H.T., Fitch J.M., Jones J.M., Schlaepfer D.D. *Trends Biochem. Sci.* 14:48-50(1989).

- [5] Klee C.B. *Biochemistry* 27:6645-6653(1988).
 [6] Smith P.D., Moss S.E. *Trends Genet.* 10:241-246(1994).
 [7] Huber R., Roemisch J., Paques E.-P. *EMBO J.* 9:3867-3874(1990).
 [8] Fiedler K., Simons K. *Trends Biochem. Sci.* 20:177-178(1995).

5

57. (arf_1) ADP-ribosylation factors family signature

ADP-ribosylation factors (ARF) [1,2,3,4] are 20 Kd GTP-binding proteins involved in protein trafficking. They may modulate vesicle budding and uncoating within the Golgi apparatus. ARF's also act as allosteric activators of cholera toxin ADP-ribosyltransferase activity. They are evolutionary conserved and present in all eukaryotes. At least six forms of ARF are present in mammals and three in budding yeast. The ARF family also includes proteins highly related to ARF's but which lack the cholera toxin cofactor activity, they are collectively known as ARL's (ARF-like). ARD1 is a 64 Kd mammalian protein of unknown biological function that contains an ARF domain at its C-terminal extremity. Proteins from the ARF family are generally included in the RAS 'superfamily' of small GTP-binding proteins [5], but they are only slightly related to the other RAS proteins. They also differ from RAS proteins in that they lack cysteine residues at their C-termini and are therefore not subject to prenylation. The ARFs are N-terminally myristoylated (the ARLs have not yet been shown to be modified in such a fashion). A conserved region in the C-terminal part of ARF's and ARL's has been selected as a signature pattern.

Consensus pattern: [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)-[GSA]-[LIVMF]-x-[WK]-[LIVM]-

Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see <[PDOC00017](#)>)

- [1] Boman A.L., Kahn R.A. *Trends Biochem. Sci.* 20:147-150(1995).
 [2] Moss J., Vaughan M. *Cell. Signal.* 4:367-399(1993).
 [3] Moss J., Vaughan M. *Prog. Nucleic Acid Res. Mol. Biol.* 45:47-65(1993).
 [4] Amor J.C., Harrison D.H., Kahn R.A., Ringe D. *Nature* 372:704-708(1994).
 [5] Valencia A., Chardin P., Wittinghofer A., Sander C. *Biochemistry* 30:4637-4648(1991).

30

(arf_2) ATP/GTP-binding site motif A (P-loop)

From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5]. There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted are listed below: - ATP synthase alpha and beta subunits (see <PDOC00137>). - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>). - Dynamins and dynamin-like proteins (see <PDOC00362>). - Guanylate kinase (see <PDOC00670>). - Thymidine kinase (see <PDOC00524>). - Thymidylate kinase (see <PDOC01034>). - Shikimate kinase (see <PDOC00868>). - Nitrogenase iron protein family (nifH/fixC) (see <PDOC00580>). - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>). - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran (see <PDOC00859>). - ADP-ribosylation factors family (see <PDOC00781>). - Bacterial dnaA protein (see <PDOC00771>). - Bacterial recA protein (see <PDOC00131>). - Bacterial recF protein (see <PDOC00539>). - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family (See <PDOC00388>). - Bacterial type II secretion system protein E (see <PDOC00567>). Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern: [AG]-x(4)-G-K-[ST]-

[1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).

- [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).
- [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).
- [4] Dever T.E., Glynias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).
- 5 [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).
- [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).
- [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).
- [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).
- 10 [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).
- [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

58. Arginase family signatures

The following enzymes have been shown [1] to be evolutionary related: - Arginase (EC 3.5.3.1), a ubiquitous enzyme which catalyzes the degradation of arginine to ornithine and urea [2]. - Arginase (EC 3.5.3.11) (arginine ureohydrolase), a prokaryotic enzyme (gene speB) that catalyzes the hydrolysis of arginine into putrescine and urea. - Formiminoglutamase (EC 3.5.3.8) (formiminoglutamate hydrolase), a prokaryotic enzyme (gene hutG) that hydrolyzes N-formimino-glutamate into glutamate and formamide. - Hypothetical proteins from methanogenic archaeobacteria. These enzymes are proteins of about 300 amino-acid residues. Three conserved regions that contain charged residues which are involved in the binding of the two manganese ions [3] can be used as signature patterns.-

Consensus pattern: [LIVMF]-G-G-x-H-x-[LIVMT]-[STAV]-x-[PAG]-x(3)-[GSTA] [H binds manganese]-

Consensus pattern: [LIVM](2)-x-[LIVMFY]-D-[AS]-H-x-D [The two D's and the H bind manganese]-

Consensus pattern: [ST]-[LIVMFY]-D-[LIVM]-D-x(3)-[PAQ]-x(3)-P-[GSA]-x(7)-G [The two D's bind manganese]

- [1] Ouzounis C., Kyripides N.C. J. Mol. Evol. 39:101-104(1994).
 [2] Jenkinson C.P., Grody W.W., Cederbaum S.D. Comp. Biochem. Physiol. 114B:107-132(196).
 [3] Kanyo Z.F., Scolnick L.R., Ash D.E., Christianson D.W. Nature 383:554-557(1996).

59. (asp) Eukaryotic and viral aspartyl proteases active site

Aspartyl proteases, also known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes [1,2,3] known to exist invertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases are: - Vertebrate gastric pepsins A and C (also known as gastricsin). - Vertebrate chymosin (rennin), involved in digestion and used for making cheese. - Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34). - Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma. - Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21). - Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases. - Yeast barrier pepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone. - Fission yeast *ssa1* which is involved in degrading or processing the mating pheromones. Most retroviruses and some plant viruses, such as badnaviruses, encode for an aspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of a polyprotein which is cleaved during the maturation process of the virus. It is generally part of the pol polyprotein and, more rarely, of the gag polyprotein. Conservation of the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases allows us to develop a single signature pattern for both groups of protease.

Consensus pattern: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-x-[LIVMFSTNC]-x-[LIVMFGTA] [D is the active site residue]

Note: these proteins belong to families A1 and A2 in the classification of peptidases [4,E1

[1] Foltmann B. Essays Biochem. 17:52-84(1981).

[2] Davies D.R. Annu. Rev. Biophys. Chem. 19:189-215(1990).

5 [3] Rao J.K.M., Erickson J.W., Wlodawer A. Biochemistry 30:4663-4671(1991).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:105-120(1995).

60. (BIRA) Biotin repressor

10 [1] Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW; Proc Natl Acad Sci USA 1992;89:9257-9261.

61. BTB/POZ domain

15 The BTB (for BR-C, ttk and bab) [1] or POZ (for Pox virus and Zinc finger)[2] domain is present near the N-terminus of a fraction of zinc finger (zf-C2H2) proteins and in proteins that contain the Kelch motif such as Kelch and a family of pox virus proteins. The BTB/POZ domain mediates homomeric dimerisation and in some instances heteromeric dimerisation [2].The structure of
20 the dimerised PLZF BTB/POZ domain has been solved and consists of a tightly intertwined homodimer. The central scaffolding of the protein is made up of a cluster of alpha-helices flanked by short beta-sheets at both the top and bottom of the molecule [3]. POZ domains from several zinc finger proteins have been shown to mediate transcriptional repression and to interact with components of histone deacetylase co-repressor complexes including N-CoR
25 and SMRT [4,5,6]. The POZ or BTB domain is also known as BR-C/Ttk or ZiN

[1] Zollman S, Godt D, Prive GG, Couderc JL, Laski FA; Proc Natl Acad Sci U S A 1994;91:10717-10721.

[2]Bardwell VJ, Treisman R; Genes Dev 1994;8:1664-1677.

30 [3] Ahmad KF, Engel CK, Prive GG; Proc Natl Acad Sci U S A 1998;95:12123-12128.

[4] Deweindt C, Albagli O, Bernardin F, Dhordain P, Quief S, Lantoine D, Kerckaert JP, Leprince D; Cell Growth Differ 1995;6:1495-1503.

[5] Huynh KD, Bardwell VJ; Oncogene 1998;17:2473-2484.

[6] Wong CW, Privalsky ML; J Biol Chem 1998;273:27695-27702.

62. (Bac GSPproteins) Bacterial type II secretion system protein D signature

5 A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1 to 5], have been found to be evolutionary related. These proteins are listed below: - The 'D' protein from the GSP operon of: *Aeromonas* (gene *exeD*); *Erwinia* (gene *outD*); *Escherichia coli* (gene *yheF*), *Klebsiella pneumoniae* (gene *pulD*); *Pseudomonas aeruginosa* (gene *xcpQ*); *Vibrio cholerae* (gene *epsD*) and *Xanthomonas campestris* (gene *xpsD*). - *comE* from *Haemophilus influenzae*, involved in competence (DNA uptake). - *pilQ* from *Pseudomonas aeruginosa*, which is essential for the formation of the pili. - *hofQ* (*hopQ*) from *Escherichia coli*. - *hrpH* from *Pseudomonas syringae*, which is involved in the secretion of a proteinaceous elicitor of the hypersensitivity response in plants. - *hrpA1* from *Xanthomonas campestris* pv. vesicatoria, which is also involved in the hypersensitivity response. - *mxlD* from *Shigella flexneri* which is involved in the secretion of the *Ipa* invasins which are necessary for penetration of intestinal epithelial cells. - *omc* from *Neisseria gonorrhoeae*. - *yssC* from *Yersinia enterocolitica* virulence plasmid *pYV*, which seems to be required for the export of the Yop virulence proteins. - The *gpIV* protein from filamentous phages such as *f1*, *ike*, or *m13*. *GpIV* is said to be involved in phage assembly and morphogenesis. These proteins all seem to start with a signal sequence and are thought to be integral proteins in the outer membrane. As a signature pattern a conserved region in the C-terminal section of these proteins has been selected

25 Consensus pattern: [GR]-[DEQKG]-[STVM]-[LIVMA](3)-[GA]-G-[LIVMFY]-x(11)-[LIVM]-P-[LIVMFYWGS]-[LIVMF]-[GSAE]-x-[LIVM]-P- [LIVMFYW](2)-x(2)-[LV]-F

[1] Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).

[2] Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G., Bunce N., Barallon R., Douglas P., Mulholland V., Stevens S., Walker S., Salmond G.P.C. Mol. Microbiol. 8:443-456(1993).

[3] Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S. Mol. Microbiol. 9:857-868(1993).

[4] Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).

[5] Genin S., Boucher C.A. Mol. Gen. Genet. 243:112-118(1994).

5 63. (Bac globin) Protozoan/cyanobacterial globins signature

Globins are heme-containing proteins involved in binding and/or transporting oxygen [1]. Almost all globins belong to a large family (see <PDOC00793>), the only exceptions are the following proteins which form a family of their own[2,3]: - Monomeric hemoglobins from the protozoan *Paramecium caudatum*, *Tetrahymena pyriformis* and *Tetrahymena thermophila*. - Cyanoglobin from the cyanobacteria *Nostoc commune*. - Globins LI637 and LI410 from the chloroplast of the alga *Chlamydomonas eugametos*. - *Mycobacterium tuberculosis* hypothetical protein MtCY48.23. These proteins contain a conserved histidine which could be involved in heme-binding. As a signature pattern, a conserved region that ends with this residue was used

15 Consensus pattern: F-[LF]-x(5)-G-[PA]-x(4)-G-[KRA]-x-[LIVM]-x(3)-H-

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

20 [2] Takagi T. Curr. Opin. Struct. Biol. 3:413-418(1993).

[3] Couture M., Chamberland H., St-Pierre B., Lafontaine J., Guertin M.; Mol. Gen. Genet. 243:185-197(1994).

25 64. Band 7 protein family signature

Mammalian band 7 protein [1] (also known as 7.2B or stomatin) is an integral membrane phosphoprotein of red blood cells thought to regulate cation conductance by interacting with other proteins of the junctional complex of the membrane skeleton. Structurally, band 7 is evolutionary related to the following proteins: - *Caenorhabditis elegans* protein mec-2 [2].

30 Mec-2 positively regulates the activity of the putative mechanosensory transduction channel. It may links the mechanosensory channel and the microtubule cytoskeleton of the touch receptor neurons. - *Caenorhabditis elegans* proteins sto-1 to sto-4. - *Caenorhabditis elegans* protein unc-1. - *Escherichia coli* hypothetical protein ybbK. - *Mycobacterium tuberculosis*

hypothetical protein MtCY277.09. - Synechocystis strain PCC 6803 hypothetical protein slr1128. - Methanococcus jannaschii hypothetical protein MJ0827. Structurally all these proteins consist of a short N-terminal domain which is followed by a transmembrane region and a variable size (from 170 to 350 residues) C-terminal domain. As a signature pattern, a conserved region located about 110 residues after the transmembrane domain was selected

Consensus pattern: R-x(2)-[LIV]-[SAN]-x(6)-[LIV]-D-x(2)-T-x(2)-W-G-[LIV]-[KRH]-[LIV]-x-[KR]-[LIV]-E-[LIV]-[KR]-

[1] Gallagher P.G., Forget B.G. *J. Biol. Chem.* 270:26358-26363(1995).

[2] Huang M., Gu G., Ferguson E.L., Chalfie M. *Nature* 378:292-295(1995).

65. Barwin domain signatures

Barwin [1] is a barley seed protein of 125 residues that binds weakly a chitin analog. It contains six cysteines involved in disulfide bonds, as shown in the following schematic representation.

```
+-----+ | ***** | *****
xxxxxxxxxxxxxxxxCxxxxxxxxCxxxxCxXXXXXXXXXXXXXXXXXXXXCx ||| +-----
-----+ +-----+ 'C': conserved cysteine involved in a disulfide bond. '*':
```

position of the patterns. Barwin is closely related to the following proteins: - Hevein, a wound-induced protein found in the latex of rubber trees. - HEL, an Arabidopsis thaliana hevein-like protein [2]. - Win1 and win2, two wound-induced proteins from potato. - Pathogenesis-related protein 4 from tobacco. Hevein and the win1/2 proteins consist of an N-terminal chitin-binding domain followed by a barwin-like C-terminal domain. Barwin and its related proteins could be involved in a defense mechanism in plants. As signature patterns, two highly conserved regions that contain some of the cysteines were selected

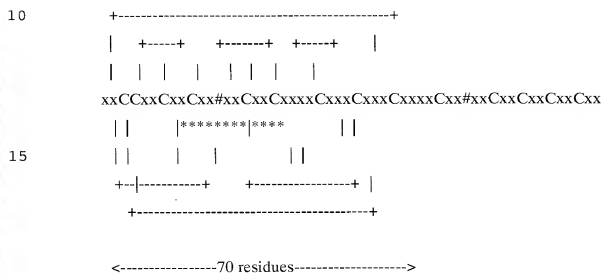
Consensus pattern: C-G-[KR]-C-L-x-V-x-N [The two C's are involved in disulfide bonds]-

Consensus pattern: V-[DN]-Y-[EQ]-F-V-[DN]-C [C is involved in a disulfide bond]-

[1] Svensson B., Svendsen I., Hoejrup P., Roepstorff P., Ludvigsen S., Poulsen F.M. *Biochemistry* 31:8767-8770(1992).

[2] Potter S., Uknes S., Lawton K., Winter A.M., Chandler D., Dimaio J., Novitzky R., Ward E., Ryals J. Mol. Plant Microbe Interact. 6:680-685(1993).

- 5 66. (Bowman-Birk leg) Bowman-Birk serine protease inhibitors family signature PROSITE cross-reference(s). The Bowman-Birk inhibitor family [1] is one of the numerous families of serine proteinase inhibitors. As it can be seen in the schematic representation, they have a duplicated structure and generally possess two distinct inhibitory sites:



'C': conserved cysteine involved in a disulfide bond.

'#': active site residue.

'*': position of the pattern.

- 25 These inhibitors are found in the seeds of all leguminous plants as well as in cereal grains. In cereals they exist in two forms, one of which is a duplication of the basic structure shown above [2]. The pattern that was developed to pick up sequences belonging to this family of inhibitors is in the central part of the domain and includes four cysteines.

30

Consensus pattern C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C [The four C's are involved in disulfide bonds] Note this pattern can be found twice in some duplicated cereal inhibitors.

[1] Laskowski M., Kato I. *Annu. Rev. Biochem.* 49:593-626(1980).

[2] Tashiro M., Hashino K., Shiozaki M., Ibuki F., Maki Z. J. *Biochem.* 102:297-306(1987).

5

67. Pathogenesis-related protein Bet v I family signature

A number of plant proteins, which all seem to be involved in pathogen defense response, are structurally related [1,2,3]. These proteins are:

10

- Bet v I, the major pollen allergen from white birch. Bet v I is the main cause of type I allergic reactions in Europe, North America and USSR.
- Aln g I, the major pollen allergen from alder.
- Api G I, the major allergen from celery.
- Car b I, the major pollen allergen from hornbeam.
- Cor a I, the major pollen allergen from hazel.
- Mal d I, the major pollen allergen from apple.
- Asparagus wound-induced protein AoPR1.
- Kidney bean pathogenesis-related proteins 1 and 2.
- Parsley pathogenesis-related proteins PR1-1 and PR1-3.
- Pea disease resistance response proteins p149, p1176 and DRRG49-C.
- Pea abscisic acid-responsive proteins ABR17 and ABR18.
- Potato pathogenesis-related proteins STH-2 and STH-21.
- Soybean stress-induced protein SAM22.

15

20

These proteins are thought to be intracellularly located. They contain from 155 to 160 amino acid residues. As a signature pattern, a conserved region located in the third quarter of these proteins has been selected

25

Consensus pattern: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]-[CS]-x(2)-K-x(4)-[FY]-

30

[1] Breiteneder H., Pettenburger K., Bito A., Valenta R., Kraft D., Rumpold H., Scheiner O., Breitenbach M. *EMBO J.* 8:1935-1938(1989).

[2] Crowell D., John M.E., Russell D., Amasino R.M. *Plant Mol. Biol.* 18:459-466(1992).

[3] Warner S.A.J., Scott R., Draper J. *Plant Mol. Biol.* 19:555-561(1992).

Consensus: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]-[CS]-x(2)-K-x(4)-[FY]-

68. bZIP transcription factors basic domain signature

The bZIP superfamily [1,2] of eukaryotic DNA-binding transcription factors groups together proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. This family is quite large, therefore only a partial list of some representative members appears here. - Transcription factor AP-1, which binds selectively to enhancer elements in the cis control regions of SV40 and metallothionein IIA. AP-1, also known as c-jun, is the cellular homolog of the avian sarcoma virus 17 (ASV17) oncogene v-jun. - Jun-B and jun-D, probable transcription factors which are highly similar to jun/AP-1. - The fos protein, a proto-oncogene that forms a non-covalent dimer with c-jun. - The fos-related proteins fra-1, and fos B. - Mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1. - Maize Opaque 2, a trans-acting transcriptional activator involved in the regulation of the production of zein proteins during endosperm. - Arabidopsis G-box binding factors GBF1 to GBF4, Parsley CPRF-1 to CPRF-3, Tobacco TAF-1 and wheat EMBP-1. All these proteins bind the G-box promoter elements of many plant genes. - Drosophila protein Giant, which represses the expression of both the *kruppel* and *knirps* segmentation gap genes. - Drosophila Box B binding factor 2 (BBF-2), a transcriptional activator that binds to fat body-specific enhancers of alcohol dehydrogenase and yolk protein genes. - Drosophila segmentation protein cap'n'collar (*gene cnc*), which is involved in head morphogenesis. - *Caenorhabditis elegans* *skn-1*, a developmental protein involved in the fate of ventral blastomeres in the early embryo. - Yeast GCN4 transcription factor, a component of the general control system that regulates the expression of amino acid-synthesizing enzymes in response to amino acid starvation, and the related *Neurospora crassa* *cpc-1* protein. - *Neurospora crassa* *cys-3* which turns on the expression of structural genes which encode sulfur-catabolic enzymes. - Yeast MET28, a transcriptional activator of sulfur amino acids metabolism. - Yeast PDR4 (or YAP1), a transcriptional activator of the genes for some oxygen detoxification enzymes. - Epstein-Barr virus trans-activator protein BZLF1.-

Consensus pattern: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK]-

[1] Hurst H.C. Protein Prof. 2:105-168(1995).[2] Ellenberger T. Curr. Opin. Struct. Biol. 4:12-21(1994).

69. Biotin-requiring enzymes attachment site

Biotin, which plays a catalytic role in some carboxyl transfer reactions, is covalently attached, via an amide bond, to a lysine residue in enzymes requiring this coenzyme [1,2,3,4]. Such enzymes are:

- Pyruvate carboxylase (EC 6.4.1.1).
- Acetyl-CoA carboxylase (EC 6.4.1.2).
- Propionyl-CoA carboxylase (EC 6.4.1.3).
- Methylcrotonyl-CoA carboxylase (EC 6.4.1.4).
- Geranoyl-CoA carboxylase (EC 6.4.1.5).
- Urea carboxylase (EC 6.3.4.6).
- Oxaloacetate decarboxylase (EC 4.1.1.3).
- Methylmalonyl-CoA decarboxylase (EC 4.1.1.41).
- Glutaconyl-CoA decarboxylase (EC 4.1.1.70).
- Methylmalonyl-CoA carboxyl-transferase (EC 2.1.3.1) (transcarboxylase).

Sequence data reveal that the region around the biocytin (biotin-lysine) residue is well conserved and can be used as a signature pattern.

Consensus pattern[GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-[SAV] [K is the biotin attachment site] Note the domain around the biotin-binding lysine residue is evolutionary related to that around the lipoyl-binding lysine residue of 2-oxo acid dehydrogenase acyltransferases

- [1] Knowles J.R. Annu. Rev. Biochem. 58:195-221(1989).
- [2] Samols D., Thronton C.G., Murtif V.L., Kumar G.K., Haase F.C., Wood H.G. J. Biol. Chem. 263:6461-6464(1988).
- [3] Goss N.H., Wood H.G. Meth. Enzymol. 107:261-278(1984).
- [4] Shenoy B.C., Xie Y., Park V.L., Kumar G.K., Beegen H., Wood H.G., Samols D. J. Biol. Chem. 267:18407-18412(1992).

2-oxo acid dehydrogenases acyltransferase component lipoyl binding site

The 2-oxo acid dehydrogenase multienzyme complexes [1,2] from bacterial and

eukaryotic sources catalyze the oxidative decarboxylation of 2-oxo acids to the corresponding acyl-CoA. The three members of this family of multienzyme complexes are:

- Pyruvate dehydrogenase complex (PDC).
- 5 - 2-oxoglutarate dehydrogenase complex (OGDC).
- Branched-chain 2-oxo acid dehydrogenase complex (BCOADC).

These three complexes share a common architecture: they are composed of multiple copies of three component enzymes - E1, E2 and E3. E1 is a thiamine pyrophosphate-dependent 2-oxo acid dehydrogenase, E2 a dihydrolipamide acyltransferase, and E3 an FAD-containing dihydrolipamide dehydrogenase.

10 E2 acyltransferases have an essential cofactor, liponic acid, which is covalently bound via an amide linkage to a lysine group. The E2 components of OGDC and BCOADC bind a single lipoyl group, while those of PDC bind either one (in yeast and in *Bacillus*), two (in mammals), or three (in *Azotobacter* and in
15 *Escherichia coli*) lipoyl groups [3].

In addition to the E2 components of the three enzymatic complexes described above, a liponic acid cofactor is also found in the following proteins:

- H-protein of the glycine cleavage system (GCS) [4]. GCS is a multienzyme complex of four protein components, which catalyzes the degradation of
20 glycine. H protein shuttles the methylamino group of glycine from the P protein to the T protein. H-protein from either prokaryotes or eukaryotes binds a single liponic group.
- Mammalian and yeast pyruvate dehydrogenase complexes differ from that of other sources, in that they contain, in small amounts, a protein of unknown
25 function - designated protein X or component X. Its sequence is closely related to that of E2 subunits and seems to bind a liponic group [5].
- Fast migrating protein (FMP) (gene *acoC*) from *Alcaligenes eutrophus* [6]. This protein is most probably a dihydrolipamide acyltransferase involved in acetoin metabolism.

30 A signature pattern was developed which allows the detection of the lipoyl-binding site.

Consensus pattern[GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-
[STAVQDN]-x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY] [K is the lipoyl-binding site] Note
the domain around the lipoyl-binding lysine residue is evolutionary related to that around the
biotin-binding lysine residue of biotin requiring enzymes

5

- [1] Yeaman S.J. Biochem. J. 257:625-632(1989).
- [2] Yeaman S.J. Trends Biochem. Sci. 11:293-296(1986).
- [3] Russel G.C., Guest J.R. Biochim. Biophys. Acta 1076:225-232(1991).
- [4] Fujiwara K., Okamura-Ikeda K., Motokawa Y. J. Biol. Chem. 261:8836-8841(1986).
- [5] Behal R.H., Browning K.S., Hall T.B., Reed L.J. Proc. Natl. Acad. Sci. U.S.A. 86:8732-8736(1989).
- [6] Priefert H., Hein S., Krueger N., Zeh K., Schmidt B., Steinbuechel A. J. Bacteriol. 173:4056-4071(1991).

10

15

70. C2 (C2 domain) Number of members: 295

Some isozymes of protein kinase C (PKC) [1,2] contain a domain, known as C2, of about 116 amino-acid residues which is located between the two copies of the C1 domain (that bind phorbol esters and diacylglycerol) (see <PDOC00379>) and the protein kinase catalytic domain (see <PDOC00100>). Regions with significant homology [3,E1] to the C2-domain have been found in the following proteins:

20

- PKC isoforms alpha, beta and gamma and Drosophila isoforms PKC1 and PKC2.
- PKC isoforms delta, epsilon and eta, Caenorhabditis elegans kin-13 and yeast PKC1 have a C2-like domain at the N-terminal extremity [4].
- Yeast cAMP dependent protein kinase SCH9 contains a C2-like domain.
- Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) (see <PDOC50007>) isoforms beta, gamma and delta as well as several non-mammalian PI-PLCs have a C2-like domain C-terminal of the catalytic domain.
- Mammalian and plants phosphatidylinositol-3-kinase have a C2-like domain in the central region of the 110 Kd catalytic subunit.
- Yeast phosphatidylserine-decarboxylase 2 (gene PSD2) contains a C2 domain in its central region.

25

30

- Cytosolic phospholipase D from plants and cytosolic phospholipase A2 have a C2-like domain at their N-terminus.

- Synaptotagmins (p65). This is a family of related synaptic vesicle proteins that bind acidic phospholipids and that may have a regulatory role in the membrane interactions during trafficking of synaptic vesicles at the active zone of the synapse. All isoforms of

- Rabphilin-3A, a synaptic protein contains two C2 domains.

- *Caenorhabditis elegans* protein unc-13 whose function is not known. Unc-13 has a C2 domain in its central part and a C2-like domain at the C-terminus.

- rasGAP and the breakpoint cluster protein bcr have a C2-domain C-terminal of a PH-domain.

- Yeast protein BUD2 (or CLA2) has a C2-domain in the central region.

- Yeast protein RSP5 and human protein NEDD-4, both proteins also contain WW domains (see <PDOC50020>).

- Perforin (see <PDOC00251>) has a C2 domain at the C-terminus. It is the only extracellular protein known to contain a C2 domain.

- Yeast hypothetical protein YML072C has a C2 domain.

- Yeast hypothetical protein YNL087W has three C2 domains.

- *Caenorhabditis elegans* hypothetical protein F37A4.7 has two C2 domains.

The C2 domain is thought to be involved in calcium-dependent phospholipid binding [5]. Since domains related to the C2 domain are also found in proteins that do not bind calcium, other putative functions for the C2 domain like e.g. binding to inositol-1,3,4,5-tetrakisphosphate have been suggested [6]. Recently, the 3D structure of the first C2 domain of synaptotagmin has been reported [7], the domain forms an eight-stranded beta sandwich. The signature pattern that has been developed for the C2 domain is located in a conserved part of that domain, the connecting loop between beta strands 2 and 3. A profile has been developed for the C2 domain that covers the total domain.

-Consensus pattern: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D-[PA]-[FY]

-Note: this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

- [1]Medline: 96367095 Extending the C2 domain family: C2s in PKCs delta, epsilon, eta and theta, phospholipases, GAPs and perforin. Ponting CP, Parker PJ; Protein Sci 1996;5:162-166.
- 5 [1] Azzi A., Boscoboinik D., Hensey C. Eur. J. Biochem. 208:547-557(1992).
- [2] Stabel S. Semin. Cancer Biol. 5:277-284(1994).
- [3] Brose N., Hofmann K.O., Hata Y., Suedhof T.C. J. Biol. Chem. 270:25273-25280(1995).
- [4] Sossin W.S., Schwartz J.H. Trends Biochem. Sci. 18:207-208(1993).
- [5] Davletov B.A., Suedhof T.C. J. Biol. Chem. 268:26386-26390(1993).
- 10 [6] Fukuda M., Aruga J., Niinobe M., Aimoto S., Mikoshiba K. J. Biol. Chem. 269:29206-29211(1994).
- [6] Sutton R.B., Davletov B.A., Berghuis A.M., Suedhof T.C., Sprang S.R. Cell 80:929-938(1995).

71. CAP (CAP protein) Number of members: 11

In budding and fission yeasts the CAP protein is a bifunctional protein whose N-terminal domain binds to adenyl cyclase, thereby enabling that enzyme to be activated by upstream regulatory signals, such as Ras. The function of the C-terminal domain is less clear, but it is required for normal cellular morphology and growth control [1]. CAP is conserved in higher eukaryotic organisms where its function is not yet clear [2].

Structurally, CAP is a protein of 474 to 551 residues which consist of two domains separated by a proline-rich hinge. Two signature patterns, one corresponding to a conserved region in the N-terminal extremity and the other to a C-terminal region have been developed.

-Consensus pattern: [LIVM](2)-x-R-L-[DE]-x(4)-R-L-E

-Consensus pattern: D-[LIVMFY]-x-E-x-[PA]-x-P-E-Q-[LIVMFY]-K

- [1] Kawamukai M., Gerst J., Field J., Riggs M., Rodgers L., Wigler M., Young D. Mol. Biol. Cell 3:167-180(1992).
- [2] Yu G., Swiston J., Young D. J. Cell Sci. 107:1671-1678(1994).

72. CAP_GLY (CAP-Gly domain)

CAP stands for cytoskeleton-associated proteins. Swiss:P39937 may be a member but has not been included. It has a weak match to the family between residues 22-67. Number of members: 24

[1]Medline: 93242656. Sequence homologies between four cytoskeleton-associated proteins. Riehemann K, Sorg C; Trends Biochem Sci 1993;18:82-83.

It has been shown [1] that some cytoskeleton-associated proteins (CAP) share the presence of a conserved, glycine-rich domain of about 42 residues, called here CAP-Gly. Proteins known to contain this domain are listed below.

- Restin (also known as cytoplasmic linker protein-170 or CLIP-170), a 160 Kd protein associated with intermediate filaments and that links endocytic vesicles to microtubules. Restin contains two copies of the CAP-Gly domain.

- Vertebrate dynactin (150 Kd dynein-associated polypeptide; DAP) and Drosophila glued, a major component of activator I, a 20S polypeptide complex that stimulates dynein-mediated vesicle transport.

- Yeast protein BIK1 which seems to be required for the formation or stabilization of microtubules during mitosis and for spindle pole body fusion during conjugation.

- Yeast protein NIP100 (NIP80).

- Human protein CKAP1/TFCB, Schizosaccharomyces pombe protein alp11 and Caenorhabditis elegans hypothetical protein F53F4.3. These proteins contain a N-terminal ubiquitin domain (see <PDOC00271>) and a C-terminal CAP-Gly domain.

- Caenorhabditis elegans hypothetical protein M01A8.2.

- Yeast hypothetical protein YNL148c.

Structurally, these proteins are made of three distinct parts: an N-terminal section that is most probably globular and contains the CAP-Gly domain, a large central region predicted to be in an alpha-helical coiled-coil conformation and, finally, a short C-terminal globular domain. The signature for the CAP-Gly domain corresponds to the first 32 residues of the domain and includes five of the six conserved glycines.

-Consensus pattern: G-x(8,10)-[FYW]-x-G-[LIVM]-x-[LIVMFY]-x(4)-G-K-[NH]-x-G-[STAR]-x(2)-G-x(2)-[LY]-F

[1] Riehemann K., Sorg C. Trends Biochem. Sci. 18:82-83(1993).

5 73. (CBD 1)

Cellulose-binding domain, fungal type

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1].

10

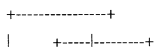
Structurally, cellulases and xylanases generally consist of a catalytic domain joined to a cellulose-binding domain (CBD) by a short linker sequence rich in proline and/or hydroxy-amino acids.

15 The CBD of a number of fungal cellulases has been shown to consist of 36 amino acid residues. Enzymes known to contain such a domain are:

- Endoglucanase I (gene egl1) from *Trichoderma reesei*.
- Endoglucanase II (gene egl2) from *Trichoderma reesei*.
- 20 - Endoglucanase V (gene egl5) from *Trichoderma reesei*.
- Exocellobiohydrolase I (gene CBHI) from *Humicola grisea*, *Neurospora crassa*, *Phanerochaete chrysosporium*, *Trichoderma reesei*, and *Trichoderma viride*.
- Exocellobiohydrolase II (gene CBHII) from *Trichoderma reesei*.
- Exocellobiohydrolase 3 (gene cel3) from *Agaricus bisporus*
- 25 - Endoglucanases B, C2, F and K from *Fusarium oxysporum*.

The CBD domain is found either at the N-terminal (Cbh-II or egl2) or at the C-terminal extremity (Cbh-I, egl1 or egl5) of these enzymes. As it is shown in the following schematic representation, there are four conserved cysteines in this type of CBD domain, all involved in disulfide bonds.

30



```

      |   |   |   |
xxxxxxx Cxxxxxxxxxxx Cxxxxx Cxxxxxxxxxxx Cx
*****

```

5 'C': conserved cysteine involved in a disulfide bond.

!*: position of the pattern.

Such a domain has also been found in a putative polysaccharide binding protein from the red alga, *Porphyra purpurea* [2]. Structurally, this protein consists of four tandem repeats of the
 10 CBD domain.

Consensus pattern C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]-x(2)-Q-C [The four C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern ALL.

15 [1] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[2] Liu Q., der Meer J.P., Reith M.E.

20 74. CBS domain. 3D Structure found as a subdomain in TIM barrel of inosine-. CBS domain web page. CBS domains are small intracellular modules mostly found in 2 or four copies within a protein. CBS domains are found in cystathionine-beta-synthase (CBS) where mutations lead to homocystinuria. Two CBS domains are found in inosine-monophosphate dehydrogenase from all species, however the CBS domains are not needed for activity. Two
 25 CBS domains are found in intracellular loops of several chloride channels. Mutations in this domain of Swiss:P35520 lead to homocystinuria.

Number of members: 414

30 [1]Medline: 97172695 The structure of a domain common to archaeobacteria and the homocystinuria disease protein. Bateman A; Trends Biochem Sci 1997;22:12-13.

[2]Medline: 96279836 Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic-acid. Sintchak MD, Fleming MA,

Futer O, Raybuck SA, Chambers SP, Caron PR, Murcko MA, Wilson KP; Cell 1996;85:921-930.

Discovery of CBS domain.

[3]Medline: 97259972 CBS domains in CIC chloride channels implicated in myotonia and nephrolithiasis (kidney stones). Ponting CP; J Mol Med 1997;75:160-163.

75. CDP-OH_P_transf (CDP-alcohol phosphatidyltransferase)

All of these members have the ability to catalyze the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond. Number of members: 32

A number of phosphatidyltransferases, which are all involved in phospholipid biosynthesis and that share the property of catalyzing the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond share a conserved sequence region [1,2]. These enzymes are:

- Ethanolaminephosphotransferase (EC 2.7.8.1) from yeast (gene EPT1).
- Diacylglycerol cholinephosphotransferase (EC 2.7.8.2) from yeast (gene CPT1).
- Phosphatidylglycerophosphate synthase (EC 2.7.8.5) (CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase) from bacteria (gene pgsA).
- Phosphatidylserine synthase (EC 2.7.8.8) (CDP-diacylglycerol--serine O-phosphatidyltransferase) from yeast (gene CHO1) and from *Bacillus subtilis* (gene pssA).
- Phosphatidylinositol synthase (EC 2.7.8.11) (CDP-diacylglycerol--inositol 3-phosphatidyltransferase) from yeast (gene PIS).

These enzymes are proteins of from 200 to 400 amino acid residues. The conserved region contains three aspartic acid residues and is located in the N-terminal section of the sequences.

-Consensus pattern: D-G-x(2)-A-R-x(8)-G-x(3)-D-x(3)-D

[1]Medline: 97075020 Two-dimensional 1H-NMR of transmembrane peptides from *Escherichia coli* phosphatidylglycerophosphate synthase in micelles. Morein S, Trouard TP, Hauksson JB, Rilfors L, Arvidson G, Lindblom G; Eur J Biochem 1996;241:489-497.

[1] Nikawa J.-I., Kodaki T., Yamashita S.

J. Biol. Chem. 262:4876-4881(1987).

[2] Hjelmstad R.H., Bell R.M.

J. Biol. Chem. 266:5094-5134(1991).

5

76. CHOD (Cholesterol oxidase) Members of the GMC oxidoreductase family. Number of members: 3

10

[1]Medline: 94032271. Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. Li J, Vrielink A, Brick P, Blow DM; Biochemistry 1993;32:11507-11515.

The following FAD flavoproteins oxidoreductases have been found [1,2] to be evolutionary related. These enzymes, which are called 'GMC oxidoreductases', are listed below.

15

- Glucose oxidase (EC 1.1.3.4) (GOX) from *Aspergillus niger*. Reaction catalyzed: glucose + oxygen -> delta-luconolactone + hydrogen peroxide.

- Methanol oxidase (EC 1.1.3.13) (MOX) from fungi. Reaction catalyzed: methanol + oxygen -> acetaldehyde + hydrogen peroxide.

20

- Choline dehydrogenase (EC 1.1.99.1) (CHD) from bacteria. Reaction catalyzed: choline + unknown acceptor -> betaine acetaldehyde + reduced acceptor.

- Glucose dehydrogenase (GLD) (EC 1.1.99.10) from *Drosophila*. Reaction catalyzed: glucose + unknown acceptor -> delta-gluconolactone + reduced acceptor.

25

- Cholesterol oxidase (CHOD) (EC 1.1.3.6) from *Brevibacterium sterolicum* and *Streptomyces* strain SA-COO. Reaction catalyzed: cholesterol + oxygen -> cholest-4-en-3-one + hydrogen peroxide.

- AlkJ [3], an alcohol dehydrogenase from *Pseudomonas oleovorans*, which converts aliphatic medium-chain-length alcohols into aldehydes. This family also includes a lyase:

- (R)-mandelonitrile lyase (EC 4.1.2.10) (hydroxynitrile lyase) from plants [4], an enzyme involved in cyanogenesis, the release of hydrogen cyanide from injured tissues.

30

These enzymes are proteins of size ranging from 556 (CHD) to 664 (MOX) amino acid residues which share a number of regions of sequence similarities. One of these regions, located in the N-terminal section, corresponds to the FAD ADP- binding domain. The function of the other conserved domains is not yet known; two of these domains have been

selected as signature patterns. The first one is located in the N-terminal section of these enzymes, about 50 residues after the ADP-binding domain, while the second one is located in the central section.

- 5 -Consensus pattern: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-[PAG]-x(5)-[DNESH]
 -Consensus pattern: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G

[1] Cavener D.R. J. Mol. Biol. 223:811-814(1992).

- 10 [2] Henikoff S., Henikoff J.G. Genomics 19:97-107(1994).

[3] van Beilen J.B., Eggink G., Enequist H., Bos R., Witholt B. Mol. Microbiol. 6:3121-3136(1992).

[4] Cheng I.P., Poulton J.E. Plant Cell Physiol. 34:1139-1143(1993).

15 77. CKS (Cyclin-dependent kinase regulatory subunit) Number of members: 11. Cyclin-dependent kinases (CDK) are protein kinases which associate with cyclins to regulate eukaryotic cell cycle progression. The most well known CDK is p34-cdc2 (CDC28 in yeast) which is required for entry into S-phase and mitosis. CDK's bind to a regulatory subunit which is essential for their biological function. This regulatory subunit is a small protein of
 20 79 to 150 residues. In yeast (gene CKS1) and in fission yeast (gene suc1) a single isoform is known, while mammals have two highly related isoforms. It has been shown [1] that these CDK regulatory subunits assemble as an hexamer which then acts as a hub for the oligomerization of six CDK catalytic subunits. The sequence of CDK regulatory subunits are
 25 highly conserved therefore, the two most conserved regions have been used as signature patterns.

-Consensus pattern: Y-S-x-[KR]-Y-x-[DE](2)-x-[FY]-E-Y-R-H-V-x-[LV]-[PT]-[KRP]

-Consensus pattern: H-x-P-E-x-H-[IV]-L-L-F-[KR]

30

[1] Parge H.E., Arvai A.S., Murtari D.J., Reed S.I., Tainer J.A. Science 262:387-395(1993).

10

15

20

[2] Reed J.C., Bidwai A.P., Glover C.V.C. J. Biol. Chem. 269:18192-18200(1994).

25

-!- The Clp protease has an active site catalytic triad. In *E. coli* Clp protease, ser-111, his-136 and asp-185 form the catalytic triad.

-1- Swiss:P48254 has lost all of these active site residues and is therefore inactive.

-!- Swiss:P42379 contains two large insertions, Swiss:P42380 contains one large insertion.

30

The endopeptidase Clp (EC 3.4.21.92) from *Escherichia coli* cleaves peptides in various proteins in a process that requires ATP hydrolysis [1,2]. Clp is a dimeric protein which consists of a proteolytic subunit (gene *clpP*) and either of two related ATP-binding regulatory subunits (genes *clpA* and *clpX*). ClpP is a serine protease which has a chymotrypsin-like activity. Its catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent

evolution. Proteases highly similar to ClpP have been found to be encoded in the genome of the chloroplast of plants and seem to be also present in other eukaryotes. The sequences around two of the residues involved in the catalytic triad (a serine and a histidine) are highly conserved and can be used as signature patterns specific to that category of proteases.

-Consensus pattern: T-x(2)-[LIVMF]-G-x-A-[SAC]-S-[MSA]-[PAG]-[STA] [S is the active site residue]

-Consensus pattern: R-x(3)-[EAP]-x(3)-[LIVMFYT]-M-[LIVM]-H-Q-P [H is the active site residue]

[1]Medline: 98050920. The structure of ClpP at 2.3 angstroms resolution suggests a model for ATP-dependent proteolysis. Wang J, Hartling JA, Flanagan JM; Cell 1997;91:447-456.

[1] Maurizi M.R., Clark W.P., Kim S.-H., Gottesman S. J. Biol. Chem. 265:12546-12552(1990).

[2] Gottesman S., Maurizi M.R. Microbiol. Rev. 56:592-621(1992).

[3] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

80. CNG_membrane (Transmembrane region cyclic Nucleotide Gated Channel)

[1]Medline: 94224763. Cyclic nucleotide-gated channels: an expanding new family of ion channels. Yau KW; Proc Natl Acad Sci USA 1994;91:3481-3483.

This family is found to the N-terminus of the cNMP_binding. Number of members: 56.

Proteins that bind cyclic nucleotides (cAMP or cGMP) share a structural domain of about 120 residues [1-3]. The best studied of these proteins is the prokaryotic catabolite gene activator (also known as the cAMP receptor protein) (gene crp) where such a domain is known to be composed of three alpha-helices and a distinctive eight-stranded, antiparallel beta-barrel structure. Such a domain is known to exist in the following proteins:

- Prokaryotic catabolite gene activator protein (CAP).

- cAMP- and cGMP-dependent protein kinases (cAPK and cGPK). Both types of kinases contains two tandem copies of the cyclic nucleotide-binding domain. The cAPK's are composed of two different subunits: a catalytic chain and a regulatory chain which contains both copies of the domain. The cGPK's are single chain enzymes that include the two copies

of the domain in their N-terminal section. The nucleotide specificity of cAPK and cGPK is due to an amino acid in the conserved region of beta-barrel 7: a threonine that is invariant in cGPK is an alanine in most cAPK.

- Vertebrate cyclic nucleotide-gated ion-channels. Two such cations channels have been fully characterized. One is found in rod cells where it plays a role in visual signal transduction. It specifically binds to cGMP leading to an opening of the channel and thereby causing a depolarization of rod photoreceptors. In olfactory epithelium a similar, cAMP-binding, channel plays a role in odorant signal transduction. There are six invariant amino acids in this domain, three of which are glycine residues that are thought to be essential for maintenance of the structural integrity of the beta-barrel. Two signature patterns have been developed for this domain. The first pattern is located within beta-barrels and 3 and contains the first two conserved Gly. The second pattern is located within beta-barrels 6 and 7 and contains the third conserved Gly as well as the three other invariant residues.

-Consensus pattern: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G

-Consensus pattern: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV]

[1] Weber I.T., Shabb J.B., Corbin J.D. Biochemistry 28:6122-6127(1989).

[2] Kaupp U.B. Trends Neurosci. 14:150-157(1991).

[3] Shabb J.B., Corbin J.D. J. Biol. Chem. 267:5723-5726(1992).

81. COX10_ctaB_cyoE (Cytochrome c oxidase assembly factor)

[1]Medline: 95191390

Biosynthesis and functional role of haem O and haem A

Mogi T, Saiki K, Anraku Y; Mol Microbiol 1994;14:391-398.

Cytochrome c oxidase is a multi subunit enzyme. The complexity of this enzyme requires assistance in building the complex.

This is carried out by the Cytochrome c oxidase assembly factor.

Number of members: 31

Cytochrome c oxidase is an oligomeric enzymatic complex which seems to require the aid of a number of proteins that either act as chaperonins to help the subunits of the enzyme to fold correctly, or assist in the assembly of the metal centers [1]. One of these subunits is known as COX10 in yeast and as ctaB [2] in aerobic prokaryotes. It is evolutionary related to cyoE protein from the Escherichia coli cytochrome O terminal oxidase complex.

These proteins probably contain [3] seven transmembrane segments. The most conserved region is located in a loop between the second and third of these segments and has been selected as a signature pattern.

-Consensus pattern: [ED]-x-D-x(2)-M-x-R-T-x(2)-R-x(4)-G

[1] Nobrega M.P., Nobrega F.G., Tzagoloff A.

J. Biol. Chem. 265:14220-14226(1990).

[2] Cao J., Hosler J., Shapleigh J., Revzin A., Ferguson-Miller S.

J. Biol. Chem. 267:24273-24278(1992).

[3] Chepuri V., Gennis R.B.

J. Biol. Chem. 265:12978-12986(1990).

82. COX3 (Cytochrome c oxidase subunit III)

This family corresponds to chains c and p.

[1]Medline: 96216288

The whole structure of the 13-subunit oxidized cytochrome c

oxidase at 2.8 Å. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S; Science 1996;272:1136-1144.
Number of members: 224

83. COX5B (Cytochrome c oxidase subunit Vb)

[1]

Medline: 96216288

The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S;

5 Science 1996;272:1136-1144.

This family consists of chains F and S

Number of members: 10

10 Cytochrome c oxidase (EC 1.9.3.1) [1] is an oligomeric enzymatic complex which is a component of the respiratory chain complex and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. In addition to the three large subunits that form the catalytic center of the enzyme complex there are, in eukaryotes, 15 a variable number of small polypeptidic subunits. One of these subunits which is known as Vb in mammals, V in slime mold and IV in yeast, binds a zinc atom. The sequence of subunit Vb is well conserved and includes three conserved cysteines that are thought to coordinate the zinc ion [2]. Two of these cysteines are clustered in the C-terminal section of the subunit; this region has been selected 20 as a signature pattern.

-Consensus pattern: [LIVM](2)-[FYW]-x(10)-C-x(2)-C-G-x(2)-[FY]-K-L [The two C's probably bind zinc]

25 [1] Capaldi R.A., Malatesta F., Darley-Usmar V.M.

Biochim. Biophys. Acta 726:135-148(1983).

[2] Rizzuto R., Sandona D., Brini M., Capaldi R.A., Bisson R.

Biochim. Biophys. Acta 1129:100-104(1991).

30

84. COesterase (Carboxylesterases)

Cholinesterase pages

The prints entry is specific to acetylcholinesterase

Number of members: 273

Higher eukaryotes have many distinct esterases. Among the different types are those which act on carboxylic esters (EC 3.1.1.-). Carboxyl-esterases have been classified into three categories (A, B and C) on the basis of differential patterns of inhibition by organophosphates. The sequence of a number of type-B carboxylesterases indicates [1,2,3] that the majority are evolutionary related. This family currently consists of the following proteins:

- Acetylcholinesterase (EC 3.1.1.7) (AChE) [E1] from vertebrates and from *Drosophila*.
- Mammalian cholinesterase II (butyryl cholinesterase) (EC 3.1.1.8).
Acetylcholinesterase and cholinesterase II are closely related enzymes that hydrolyze choline esters [4].
- Mammalian liver microsomal carboxylesterases (EC 3.1.1.1).
- *Drosophila* esterase 6, produced in the anterior ejaculatory duct of the male insect reproductive system where it plays an important role in its reproductive biology.
- *Drosophila* esterase P.
- *Culex pipiens* (mosquito) esterases B1 and B2.
- *Myzus persicae* (peach-potato aphid) esterases E4 and FE4.
- Mammalian bile-salt-activated lipase (BAL) [5], a multifunctional lipase which catalyzes fat and vitamin absorption. It is activated by bile salts in infant intestine where it helps to digest milk fats.
- Insect juvenile hormone esterase (JH esterase) (EC 3.1.1.59).
- Lipases (EC 3.1.1.3) from the fungi *Geotrichum candidum* and *Candida rugosa*.
- *Caenorhabditis* gut esterase (gene ges-1).
- Duck fatty acyl-CoA hydrolase, medium chain (EC 3.1.2.14), an enzyme that may be associated with peroxisome proliferation and may play a role in the production of 3-hydroxy fatty acid diester pheromones.
- Membrane enclosed crystal proteins from slime mold. These proteins are, most probably esterases; the vesicles where they are found have therefore

been termed esterosomes.

So far two bacterial proteins have been found to belong to this family:

- 5 - Phenmedipham hydrolase (phenylcarbamate hydrolase), an *Arthrobacter* oxidans plasmid-encoded enzyme (gene *pcd*) that degrades the phenylcarbamate herbicides phenmedipham and desmedipham by hydrolyzing their central carbamate linkages.
- Para-nitrobenzyl esterase from *Bacillus subtilis* (gene *pnbA*).

10

The following proteins, while having lost their catalytic activity, contain a domain evolutionary related to that of carboxylesterases type-B:

- 15 - Thyroglobulin (TG), a glycoprotein specific to the thyroid gland, which is the precursor of the iodinated thyroid hormones thyroxine (T4) and triiodo thyronine (T3).
- *Drosophila* protein neuractin (gene *nrt*) which may mediate or modulate cell adhesion between embryonic cells during development.
- *Drosophila* protein glutactin (gene *glt*), whose function is not known.

20

As is the case for lipases and serine proteases, the catalytic apparatus of esterases involves three residues (catalytic triad): a serine, a glutamate or aspartate and a histidine. The sequence around the active site serine is well conserved and can be used as a signature pattern. A conserved region located in the N-terminal section containing a cysteine involved in a disulfide bond

25 has been selected as a second signature pattern.

-Consensus pattern: F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G[S] is the active site residue]

- 30 -Consensus pattern: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR] [C is involved in a disulfide bond]

[1] Myers M., Richmond R.C., Oakeshott J.G. Mol. Biol. Evol. 5:113-119(1988).

[2] Krejci E., Duval N., Chatonnet A., Vincens P., Massoulie J. Proc. Natl. Acad. Sci. U.S.A. 88:6647-6651(1991).

[3] Cygler M., Schrag J.D., Sussman J.L., Harel M., Silman I. Gentry M.K., Doctor B.P. Protein Sci. 2:366-382(1993).

5 [4] Lockridge O. BioEssays 9:125-128(1988).

[5] Wang C.-S., Hartsuck J.A. Biochim. Biophys. Acta 1166:1-19(1993).

85. CPSase_L_chain (Carbamoyl-phosphate synthase (CPSase))

10 [1]

Medline: 94347758

Three-dimensional structure of the biotin carboxylase subunit.
of acetyl-CoA carboxylase.

Waldrop GL, Rayment I, Holden HM;

15 Biochemistry 1994;33:10249-10256.

[1]

Medline: 90285162

Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and
evolution of the CPS domain of the Syrian hamster multifunctional
protein CAD.

Simmer JP, Kelly RE, Rinker AG Jr, Scully JL, Evans DR;

Biol Chem 1990;265:10395-10402.

Carbamoyl-phosphate synthase catalyzes the ATP-dependent synthesis of
carbamyl-phosphate from glutamine or ammonia and bicarbonate. This
25 important enzyme initiates both the urea cycle and the biosynthesis
of arginine and/or pyrimidines [2].

The carbamoyl-phosphate synthase (CPS) enzyme in prokaryotes is a
heterodimer of a small and large chain. The small chain promotes
the hydrolysis of glutamine to ammonia, which is used by the large
30 chain to synthesize carbamoyl phosphate. See CPSase_sm_chain.

The small chain has a GATase domain in the carboxyl terminus.

See GATase.

Number of members: 181

Carbamoyl-phosphate synthase (CPSase) catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine (EC 6.3.5.5) or ammonia (EC 6.3.4.16) and bicarbonate [1]. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and pyrimidines.

Glutamine-dependent CPSase (CPSase II) is involved in the biosynthesis of pyrimidines and purines. In bacteria such as *Escherichia coli*, a single enzyme is involved in both biosynthetic pathways while other bacteria have separate enzymes. The bacterial enzymes are formed of two subunits. A small chain (gene *carA*) that provides glutamine amidotransferase activity (GATase) necessary for removal of the ammonia group from glutamine, and a large chain (gene *carB*) that provides CPSase activity. Such a structure is also present in fungi for arginine biosynthesis (genes *CPA1* and *CPA2*). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme - called *URA2* in yeast, rudimentary in *Drosophila* and *CAD* in mammals [2]. The CPSase domain is located between an N-terminal GATase domain and the C-terminal part which encompasses the dihydroorotase and aspartate transcarbamylase activities.

Ammonia-dependent CPSase (CPSase I) is involved in the urea cycle in ureolytic vertebrates; it is a monofunctional protein located in the mitochondrial matrix.

The CPSase domain is typically 120 Kd in size and has arisen from the duplication of an ancestral subdomain of about 500 amino acids. Each subdomain independently binds to ATP and it is suggested that the two homologous halves act separately, one to catalyze the phosphorylation of bicarbonate to carboxy phosphate and the other that of carbamate to carbamyl phosphate.

The CPSase subdomain is also present in a single copy in the biotin-dependent enzymes acetyl-CoA carboxylase (EC 6.4.1.2) (ACC), propionyl-CoA carboxylase (EC 6.4.1.3) (PCCase), pyruvate carboxylase (EC 6.4.1.1) (PC) and urea

carboxylase (EC 6.3.4.6).

Two conserved regions which are probably important for binding ATP and/or catalytic activity have been selected as signatures for the subdomain.

-Consensus pattern: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG]

-Consensus pattern: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC]

[1] Simmer J.P., Kelly R.E., Rinker A.G. Jr., Scully J.L., Evans D.R.
J. Biol. Chem. 265:10395-10402(1990).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B.
BioEssays 15:157-164(1993).

86. CPSase_sm_chain (Carbamoyl-phosphate synthase small chain, CPSase domain)
[1]

Medline: 90285162

Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD.

Simmer JP, Kelly RE, Rinker AG Jr, Scully JL, Evans DR;
Biol Chem 1990;265:10395-10402.

The carbamoyl-phosphate synthase domain is in the amino terminus of protein.

Carbamoyl-phosphate synthase catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine or ammonia and bicarbonate. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and/or pyrimidines [1].

The carbamoyl-phosphate synthase (CPS) enzyme in prokaryotes is a heterodimer of a small and large chain. The small chain promotes the hydrolysis of glutamine to ammonia, which is used by the large chain to synthesize carbamoyl phosphate. See CPSase_L_chain.

Number of members: 46

- 5 Carbamoyl-phosphate synthase (CPSase) catalyzes the ATP-dependent synthesis of carbamyl-phosphate from glutamine (EC 6.3.5.5) or ammonia (EC 6.3.4.16) and bicarbonate [1]. This important enzyme initiates both the urea cycle and the biosynthesis of arginine and pyrimidines.
- 10 Glutamine-dependent CPSase (CPSase II) is involved in the biosynthesis of pyrimidines and purines. In bacteria such as *Escherichia coli*, a single enzyme is involved in both biosynthetic pathways while other bacteria have separate enzymes. The bacterial enzymes are formed of two subunits. A small chain (gene *carA*) that provides glutamine amidotransferase activity (GATase) necessary for
- 15 removal of the ammonia group from glutamine, and a large chain (gene *carB*) that provides CPSase activity. Such a structure is also present in fungi for arginine biosynthesis (genes *CPA1* and *CPA2*). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme - called URA2 in yeast, rudimentary in *Drosophila* and
- 20 CAD in mammals [2]. The CPSase domain is located between an N-terminal GATase domain and the C-terminal part which encompass the dihydroorotase and aspartate transcarbamylase activities.
- Ammonia-dependent CPSase (CPSase I) is involved in the urea cycle in ureolytic
- 25 vertebrates; it is a monofunctional protein located in the mitochondrial matrix.
- The CPSase domain is typically 120 Kd in size and has arisen from the duplication of an ancestral subdomain of about 500 amino acids. Each subdomain
- 30 independently binds to ATP and it is suggested that the two homologous halves act separately, one to catalyze the phosphorylation of bicarbonate to carboxy phosphate and the other that of carbamate to carbamyl phosphate.

The CPSase subdomain is also present in a single copy in the biotin-dependent enzymes acetyl-CoA carboxylase (EC 6.4.1.2) (ACC), propionyl-CoA carboxylase (EC 6.4.1.3) (PCCase), pyruvate carboxylase (EC 6.4.1.1) (PC) and urea carboxylase (EC 6.3.4.6).

5

Two conserved regions which are probably important for binding ATP and/or catalytic activity have been selected as signatures for the subdomain.

10

-Consensus pattern: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-
[SG]-G-x-[AG]

-Consensus pattern: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC]

[1] Simmer J.P., Kelly R.E., Rinker A.G. Jr., Scully J.L., Evans D.R.
J. Biol. Chem. 265:10395-10402(1990).

15

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B.
BioEssays 15:157-164(1993).

87. CRAL_TRIO (CRAL/TRIO domain)

20

[1]

Medline: 98121119

Crystal structure of the *Saccharomyces cerevisiae* phosphatidyl-
inositol-transfer protein.

Sha B, Phillips SE, Bankaitis VA, Luo M;

25

Nature 1998;391:506-510.

The original profile has been extended to include the carboxyl
domain from the known structure of Sec14. Swiss:P10911 has not
been included in the Pfam family because it does not appear to
contain a complete structural domain.

30

Number of members: 39

88. CSD ('Cold-shock' DNA-binding domain)

[1]

Medline: 94255482

Crystal structure of CspA, the major cold shock protein of *Escherichia coli*.

- 5 Schindelin H, Jiang W, Inouye M, Heinemann U;
Proc Natl Acad Sci U S A 1994;91:5119-5123.

Number of members: 121

- 10 A conserved domain of about 70 amino acids has been found in prokaryotic and eukaryotic DNA-binding proteins [1,2,3,E1]. This domain, which is known as the 'cold-shock domain' (CSD) is present in the proteins listed below.

- *Escherichia coli* protein CS7.4 (gene *cspA*) which is induced in response to low temperature (cold-shock protein) and which binds to and stimulates the transcription of the CCAAT-containing promoters of the *HN-S* protein and of *gyrA*.
- Mammalian Y box binding protein 1 (YB1). A protein that binds to the CCAAT-containing Y box of mammalian HLA class II genes.
- *Xenopus* Y box binding proteins -1 and -2 (Y1 and Y2). Proteins that bind to the CCAAT-containing Y box of *Xenopus hsp70* genes.
- *Xenopus* B box binding protein (YB3). YB3 binds the B box promoter element of genes transcribed by RNA polymerase III.
- Enhancer factor I subunit A (EFI-A) (dbpB). A protein that also binds to CCAAT-motif in various gene promoters.
- DbpA, a Human DNA-binding protein of unknown specificity.
- *Bacillus subtilis* cold-shock proteins *cspB* and *cspC*.
- *Streptomyces clavuligerus* protein SC 7.0.
- *Escherichia coli* proteins *cspB*, *cspC*, *cspD*, *cspE* and *cspF*.
- Unr, a mammalian gene encoded upstream of the N-ras gene. Unr contains nine repeats that are similar to the CSD domain. The function of Unr is not yet known but it could be a multivalent DNA-binding protein.

As a signature pattern for the CSD domain, its most conserved

region which is located in its N-terminal section has been selected. It must be noted that the beginning of this region is highly similar [4] to the RNP-1 RNA-binding motif.

-Consensus pattern: [FY]-G-F-I-x(6,7)-[DER]-[LIVM]-F-x-H-x-[STKR]-x-[LIVMFY]

[1] Doniger J., Landsman D., Gonda M.A., Wistow G.

New Biol. 4:389-395(1992).

[2] Wistow G.

Nature 344:823-824(1990).

[3] Jones P.G., Inouye M.

Mol. Microbiol. 11:811-818(1994).

[4] Landsman D.

Nucleic Acids Res. 20:2861-2864(1992).

89. CTF NFI (CTF/NF-I family)

Number of members: 45

Nuclear factor I (NF-I) or CCAAT box-binding transcription factor (CTF) [1,2]

(also known as TGGCA-binding proteins) are a family of vertebrate nuclear proteins which recognize and bind, as dimers, the palindromic DNA sequence 5'-TGGCANNNTGCCA-3'. CTF/NF-I binding sites are present in viral and cellular promoters and in the origin of DNA replication of Adenovirus type 2.

The CTF/NF-I proteins were first identified as nuclear factor I, a collection of proteins that activate the replication of several Adenovirus serotypes (together with NF-II and NF-III) [3]. The family of proteins was also identified as the CTF transcription factors, before the NFI and CTF families were found to be identical [4]. The CTF/NF-I proteins are individually capable of activating transcription and DNA replication. The CTF/NF-I family name has also been dubbed as NFI, NF-I or NFI.

In a given species, there are a large number of different CTF/NF-I proteins.

The multiplicity of CTF/NF-I is known to be generated both by alternative splicing and by the occurrence of four different genes. The known forms of NF-I genes have been classified as:

- 5 - The CTF-like factors subfamily (prototype form: CTF-1) [4]
- The NFI-X proteins.
- The NFI-A proteins.
- The NFI-B proteins.

- 10 So far, all CTF/NF-I family members appear to have similar transcription and replication activities.

CTF/NF-1 proteins contains 400 to 600 amino acids. The N-terminal 200 amino-acid sequence, almost perfectly conserved in all species and genes sequenced, mediates site-specific DNA recognition, protein dimerization and Adenovirus DNA replication. The C-terminal 100 amino acids contain the transcriptional activation domain. This activation domain is the target of gene expression regulatory pathways elicited by growth factors and it interacts with basal transcription factors and with histone H3 [6].

A perfectly conserved, highly charged 12 residue peptide located in the N-terminal part of CTF/NF-I has been selected as a specific signature for this family of proteins.

-Consensus pattern: R-K-R-K-Y-F-K-K-H-E-K-R

[1] Mermod N., O'Neill E.A., Kelly T.J., Tjian R.
Cell 58:741-753(1989).

[2] Rupp R.A.W., Kruse U., Multhaup G., Goebel U., Beyreuther K.,
Sippel A.E.

Nucleic Acids Res. 18:2607-2616(1990).

[3] Nagata K., Guggenheimer R.A., Enomoto T., Lichy J.H., Hurwitz J.
Proc. Natl. Acad. Sci. U.S.A. 79:6438-6442(1982).

[4] Santoro C., Mermod N., Andrews P.C., Tjian R.

Nature 334:2118-2224(1988).

[5] Gil G., Smith J.R., Goldstein J.L., Slaughter C.A., Orth K., Brown M.S., Osborne T.F.

Proc. Natl. Acad. Sci. U.S.A 85:8963-8967(1988).

5 [6] Alevizopoulos A., Dusserre Y., Tsai-Pflugfelder M., von der Weid T., Wahli W., Mermod N.

Genes Dev. 9:3051-3066(1995).

10 90. Calsequestrin (Calsequestrin)

Number of members: 13

Calsequestrin is a moderate-affinity, high-capacity calcium-binding protein of cardiac and skeletal muscle [1], where it is located in the luminal space of the sarcoplasmic reticulum terminal cisternae. Calsequestrin acts as a calcium buffer and plays an important role in the muscle excitation-contraction coupling. It is a highly acidic protein of about 400 amino acid residues that binds more than 40 moles of calcium per mole of protein. There are at least two different forms of calsequestrin: one which is expressed in cardiac muscles and another in skeletal muscles. Both forms have highly similar sequences.

Two signature sequences have been developed. The first corresponds to the N-terminus of the mature protein, the second is located just in front of the C-terminus of the protein which is composed of a highly acidic tail of variable length.

-Consensus pattern: [EQ]-[DE]-G-L-[DN]-F-P-x-Y-D-G-x-D-R-V

-Consensus pattern: [DE]-L-E-D-W-[LIVM]-E-D-V-L-x-G-x-[LIVM]-N-T-E-D-D-D

[1] Treves S., Vilsen B., Chiozzi P., Andersen J.P., Zorzato F.

Biochem. J. 283:767-772(1992).

91. Carboxyl_trans (Carboxyl transferase domain)

[1]

Medline: 93374821

- 5 Primary structure of the monomer of the 12S subunit of transcarboxylase as deduced from DNA and characterization of the product expressed in *Escherichia coli*.
 Thornton CG, Kumar GK, Haase FC, Phillips NF, Woo SB, Park VM, Magner WJ, Shenoy BC, Wood HG, Samols D;
 10 J Bacteriol 1993;175:5301-5308.

[2]

Medline: 93358891

Molecular evolution of biotin-dependent carboxylases.

- 15 Toh H, Kondo H, Tanabe T;
 Eur J Biochem 1993;215:687-696.
 All of the members in this family are biotin dependent carboxylases.
 The carboxyl transferase domain carries out the following reaction;
 transcarboxylation from biotin to an acceptor molecule. There are
 two recognised types of carboxyl transferase. One of them uses acyl-CoA
 20 and the other uses 2-oxo acid as the acceptor molecule of carbon dioxide.
 All of the members in this family utilise acyl-CoA as the acceptor
 molecule.

Number of members: 47

25

92. Chal_stil_synt (Chalcone and stilbene synthases)

Number of members: 146

- 30 Chalcone synthases (CHS) (EC 2.3.1.74) and stilbene synthases (STS) (formerly known as resveratrol synthases) are related plant enzymes [1]. CHS is an important enzyme in flavanoid biosynthesis and STS a key enzyme in stilbene-type phytoalexin biosynthesis. Both enzymes catalyze the addition of three molecules of malonyl-CoA to a starter CoA ester (a typical example is

4-coumaroyl-CoA), producing either a chalcone (with CHS) or stilbene (with STS).

These enzymes are proteins of about 390 amino-acid residues. A conserved cysteine residue, located in the central section of these proteins, has been shown [2] to be essential for the catalytic activity of both enzymes and probably represents the binding site for the 4-coumaroyl-CoA group. The region around this active site residue is well conserved and can be used as a signature pattern.

In addition to the plant enzymes, this family also includes *Bacillus subtilis* bcsA.

-Consensus pattern: R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GA]-G-[GA]-[STAV]-x-[LIVMF]-[RA] [C is the active site residue]

[1] Schroeder J., Schroeder G.

Z. Naturforsch. 45C:1-8(1990).

[2] Lanz T., Tropf S., Marner F.-J., Schroeder J., Schroeder G.

J. Biol. Chem. 266:9971-9976(1991).

93. Chorismate_synt (Chorismate synthase)

Number of members: 19

Chorismate synthase (EC 4.6.1.4) catalyzes the last of the seven steps in the shikimate pathway which is used in prokaryotes, fungi and plants for the biosynthesis of aromatic amino acids. It catalyzes the 1,4-trans elimination of the phosphate group from 5-enolpyruvylshikimate-3-phosphate (EPSP) to form chorismate which can then be used in phenylalanine, tyrosine or tryptophan biosynthesis. Chorismate synthase requires the presence of a reduced flavin mononucleotide (FMN_{H2} or FAD_{H2}) for its activity.

Chorismate synthase from various sources shows [1,2] a high degree of sequence conservation. It is a protein of about 360 to 400 amino-acid residues.

Three signature patterns have been developed from conserved regions rich in basic residues (mostly arginines). The first is in the N-terminal section, the

second is central and the third is C-terminal.

-Consensus pattern: G-E-S-H-[GC]-x(2)-[LIVM]-[GTV]-x-[LIVM](2)-[DE]-G-x-[PV]

-Consensus pattern: [GE]-R-[SA](2)-[SAG]-R-[EV]-[ST]-x(2)-[RH]-V-x(2)-G

-Consensus pattern: R-[SH]-D-[PSV]-[CSAV]-x(4)-[GAI]-x-[IVGSP]-[LIVM]-x-E-[STAH]-[LIVM]

[1] Schaller A., Schmid J., Leibinger U., Amrhein N.
J. Biol. Chem. 266:21434-21438(1991).

[2] Jones D.G.L., Reusser U., Braus G.H.
Mol. Microbiol. 5:2143-2152(1991).

94. Clat_adaptor_s (Clathrin adaptor complex small chain)

Number of members: 21

Clathrin coated vesicles (CCV) mediate intracellular membrane traffic such as receptor mediated endocytosis. In addition to clathrin, the CCV are composed of a number of other components including oligomeric complexes which are known as adaptor or clathrin assembly proteins (AP) complexes [1]. The adaptor complexes are believed to interact with the cytoplasmic tails of membrane proteins, leading to their selection and concentration. In mammals two type of adaptor complexes are known: AP-1 which is associated with the Golgi complex and AP-2 which is associated with the plasma membrane. Both AP-1 and AP-2 are heterotetramers that consist of two large chains - the adaptins - (gamma and beta' in AP-1; alpha and beta in AP-2); a medium chain (AP47 in AP-1; AP50 in AP-2) and a small chain (AP19 in AP-1; AP17 in AP-2).

The small chains of AP-1 and AP-2 are evolutionary related proteins of about 18 Kd. Homologs of AP17 and AP19 have also been found in yeast (genes APS1/YAP19 and APS2/YAP17) [2,3,4]. AP17 and AP19 are also related to the zeta-chain [5] of coatomer (zeta-cop), a cytosolic protein complex that reversibly associates with Golgi membranes to form vesicles that mediate biosynthetic protein transport from the endoplasmic reticulum, via the Golgi up to the trans Golgi network.

A conserved region in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM](2)-Y-[KR]-x(4)-L-Y-F

[1] Pearse B.M., Robinson M.S.

Annu. Rev. Cell Biol. 6:151-171(1990).

[2] Kirchhausen T., Davis A.C., Frucht S., O'Brine Greco B., Payne G.S., Tubb B.

J. Biol. Chem. 266:11153-11157(1991).

[3] Nakai M., Takada T., Endo T.

Biochim. Biophys. Acta 1174:282-284(1993).

[4] Phan H.L., Finlay J.A., Chu D.S., Tan P.K., Kirchhausen T., Payne G.S.

EMBO J. 13:1706-1717(1994).

[5] Kuge O., Hara-Kuge S., Orci L., Ravazzola M., Amherdt M., Tanigawa G., Wieland F.T., Rothman J.E.

J. Cell Biol. 123:1727-1734(1993).

95. Clathrin_lg_ch (Clathrin light chain.)

Number of members: 8

Clathrin [1,2] is the major coat-forming protein that encloses vesicles such as coated pits and forms cell surface patches involved in membrane traffic within eukaryotic cells. The clathrin coats (called triskelions) are composed

of three heavy chains (180 Kd) and three light chains (23 to 27 Kd).

The clathrin light chains [3], which may help to properly orient the assembly and disassembly of the clathrin coats, bind non-covalently to the heavy chain, they also bind calcium and interact with the hsc70 uncoating ATPase.

- In higher eukaryotes two genes code for distinct but related light chains: LC(a) and LC(b). Each of the two genes can yield, by tissue-specific alternative splicing, two separate forms which differ by the insertion of a sequence of respectively thirty or eighteen residues. There is, in the N-terminal part of the clathrin light chains a domain of twenty one amino acid residues which is perfectly conserved in LC(a) and LC(b).

- In yeast there is a single light chain (gene CLC1) whose sequence is only distantly related to that of higher eukaryotes.

Two signature patterns have been developed for clathrin light chains. The first pattern is a heptapeptide from the center of the conserved N-terminal region of eukaryotic light chains; the second pattern is derived from a positively charged region located in the C-terminal extremity of all known clathrin light chains.

-Consensus pattern: F-L-A-Q-Q-E-S

[1] Keen J.H.

Annu. Rev. Biochem. 59:415-438(1990).

[2] Brodsky F.M.

Science 242:1396-1402(1988).

[3] Brodsky F.M., Hill B.L., Acton S.L., Naethke I., Wong D.H.,

Ponnambalam S., Parham P.

Trends Biochem. Sci. 16:208-213(1991).

96. (Clathrin repeat) 7-fold repeat in Clathrin and VPS

[1]

Folding and trimerization of clathrin subunits at the triskelion hub.

Nathke IS, Heuser J, Lupas A, Stock J, Turck CW, Brodsky FM;
Cell 1992;68:899-910. [2]

Clathrin heavy chain: molecular cloning and complete primary structure.

Kirchhausen T, Harrison SC, Chow EP, Mattaliano RJ, Ramachandran KL, Smart J, Brosius J;

15 Proc Natl Acad Sci U S A 1987;84:8805-8809.

97. Collagen (Collagen triple helix repeat (20 copies))

[1] Medline: 94059583

20 New members of the collagen superfamily

Mayne R, Brewton RG:

Curr Opin Cell Biol 1993;5:883-890.

Scurvy is associated with collagens.

Members of this family belong to the collagen superfamily [1].

25 Collagens are generally extracellular structural proteins

involved in formation of connective tissue structure.

The alignment contains 20 copies of the G-X-Y repeat that

forms a triple helix. The first position of the repeat is

glycine, the second and third positions can be any residue

30 but are frequently proline and hydroxyproline. Collagens

are post translationally modified by proline hydroxylase

to form the hydroxyproline residues. Defective

hydroxylation is the cause of scurvy.

Some members of the collagen superfamily are not involved in connective tissue structure but share the same triple helical structure.

Number of members: 2125

98. Coprogen_oxidas (Coproporphyrinogen III oxidase)

Number of members: 12

Coproporphyrinogen III oxidase (EC 1.3.3.3) (coproporphyrinogenase) [1,2]

catalyzes the oxidative decarboxylation of coproporphyrinogen III into protoporphyrinogen IX, a common step in the pathway for the biosynthesis of porphyrins such as heme, chlorophyll or cobalamin.

Coproporphyrinogen III oxidase is an enzyme that requires iron for its activity. A cysteine seems to be important for the catalytic mechanism [3]. Sequences from a variety of eukaryotic and prokaryotic sources show that this enzyme has been evolutionarily conserved. A highly conserved region in the central part of the sequence has been selected as a signature pattern. This region contains the only conserved cysteine and is rich in charged amino acids.

-Consensus pattern: K-x-W-C-x(2)-[FYH](3)-[LIVM]-x-H-R-x-E-x-R-G-[LIVM]-G-G-[LIVM]-F-F-D

[1] Xu K., Elliott T.

J. Bacteriol. 175:4990-4999(1993).

[2] Kohno H., Furukawa T., Yoshinaga T., Tokunaga R., Taketani S.

J. Biol. Chem. 268:21359-21363(1993).

[3] Camadro J.M., Chambon H., Jolles J., Labbe P.

Eur. J. Biochem. 156:579-587(1986).

[4] Xu K., Elliott T.

J. Bacteriol. 176:3196-3203(1994).

99. Corona nucleoca (Coronavirus nucleocapsid protein)

[1]

Medline: 98087828

5 Identification of a specific interaction between the
coronavirus mouse hepatitis virus A59 nucleocapsid protein
and packaging signal.

Molenkamp R, Spaan WJ;

Virology 1997;239:78-86.

10 Number of members: 44

100. Cu-oxidase (Multicopper oxidase)

[1]

15 Medline: 90126844

The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin.

Modelling and structural relationships.

Messerschmidt A, Huber R:

Eur J Biochem 1990;187:341-352.

20 Number of members: 150

Multicopper oxidases [1,2] are enzymes that possess three spectroscopically different copper centers. These centers are called: type 1 (or blue), type 2 (or normal) and type 3 (or coupled binuclear). The enzymes that belong to this family are:

- Laccase (EC 1.10.3.2) (urishiol oxidase), an enzyme found in fungi and plants, which oxidizes many different types of phenols and diamines.

- Ascorbate oxidase (EC 1.10.3.3), a higher plant enzyme.

30 - Ceruloplasmin (EC 1.16.3.1) (ferroxidase), a protein found in the serum of mammals and birds, which oxidizes a great variety of inorganic and organic substances. Structurally ceruloplasmin exhibits internal sequence homology, and seem to have evolved from the triplication of a copper-binding domain

similar to that found in laccase and ascorbate oxidase.

In addition to the above enzymes there are a number of proteins which, on the basis of sequence similarities, can be said to belong to this family. These proteins are:

- Copper resistance protein A (copA) from a plasmid in *Pseudomonas syringae*. This protein seems to be involved in the resistance of the microbial host to copper.
- Blood coagulation factor V (Fa V).
- Blood coagulation factor VIII (Fa VIII) [E1].
- Yeast FET3 [3], which is required for ferrous iron uptake.
- Yeast hypothetical protein YFL041w and SpAC1F7.08, the fission yeast homolog.

Factors V and VIII act as cofactors in blood coagulation and are structurally similar [4]. Their sequence consists of a triplicated A domain, a B domain and a duplicated C domain; in the following order: A-A-B-A-C-C. The A-type domain is related to the multicopper oxidases.

Two signature patterns have been developed for these proteins. Both patterns are derived from the same region, which in ascorbate oxidase, laccase, in the third domain of ceruloplasmin, and in copA, contains five residues that are known to be involved in the binding of copper centers. The first pattern does not make any assumption on the presence of copper-binding residues and thus can detect domains that have lost the ability to bind copper (such as those in Fa V and Fa VIII), while the second pattern is specific to copper-binding domains.

- Consensus pattern: G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x(8)-G-[LM]-x(3)-[LIVMFYW]
- Consensus pattern: H-C-H-x(3)-H-x(3)-[AG]-[LM]
- [The first two H's are copper type 3 binding residues]
- [The C, the 3rd H, and L or M are copper type 1 ligands]

101. Cullin (Cullin family)

Number of members: 24

5

The following proteins are collectively termed cullins [1]:

10

- *Caenorhabditis elegans* cul-1 (or lin-19), a protein required for developmentally programmed transitions from the G1 phase of the cell cycle to the G0 phase or the apoptotic pathway.

- *Caenorhabditis elegans* cul-2, cul-3, cul-4 (F45E12.3), cul-5 (ZK856.1) and cul-6 (K08E7.7).

- Mammalian CUL1, CUL2, CUL3, CUL4A and CUL4B.

15

- Mammalian vasopressin-activated calcium-mobilizing receptor (VACM-1), a kidney-specific protein thought to form a cell surface receptor [2] but which does not have any structural hallmarks of a receptor.

- *Drosophila* lin19.

- Yeast CDC53 [3], which acts in concert with CDC4 and UBC3 (CDC34) to control the G1-to-S phase transition.

20

- Yeast hypothetical protein YGR003w.

- Fission yeast hypothetical protein SpAC24H6.03.

The cullins are hydrophilic proteins of 740 to 815 amino acids. The C-terminal extremity is the most conserved part of these proteins. A

25

signature pattern has been developed from that region.

-Consensus pattern: [LIV]-K-x(2)-[LIV]-x(2)-L-I-[DEQ]-[KRHNQ]-x-Y-[LIVM]-x-R-x(6,7)-[FY]-x-Y-x-[SA]>

30

[1] Kipreos E.T., Lander L.E., Wing J.P., He W.W., Hedgecock E.M.
Cell 85:829-839(1996).

[2] Burnatowska-Hledin M.A., Spielman W.S., Smith W.L., Shi P., Meyer J.M.,
Dewitt D.L.

Am. J. Physiol. 268:f1198-F1210(1995).

[3] Mathias N., Johnson S.L., Winey M., Adams A.E., Goetsch L., Pringle J.R.,

Byers B., Goebl M.G.

Mol. Cell. Biol. 16:6634-6643(1996).

5

102. (Cu_amine_oxid)

Copper amine oxidase signatures

Amine oxidases (AO) [1] are enzymes that catalyze the oxidation of a wide range of biogenic amines including many neurotransmitters, histamine and xenobiotic amines. There are two classes of amine oxidases: flavin-containing (EC 1.4.3.4) and copper-containing (EC 1.4.3.6).

10

Copper-containing AO is found in bacteria, fungi, plants and animals, it is an homodimeric enzyme that binds one copper ion per subunit as well as a 2,4,5- trihydroxyphenylalanine quinone (or topaquinone) (TPQ) cofactor. This cofactor is derived from a tyrosine residue.

15

Two signature patterns were derived for copper AO, the first one contains the tyrosine which give rises to the TPQ cofactor while the second one contains one of the three histidines that bind the copper atom [2].

20

Consensus pattern[LIVM]-[LIVMA]-[LIVMF]-x(4)-[ST]-x(2)-N-Y-[DE]-[YN] [The first Y gives rises to TPQ] Sequences known to belong to this class detected by the patternALL.

Consensus patternT-x-[GS]-x(2)-H-[LIVMF]-x(3)-E-[DE]-x-P [H is a copper ligand]

25

Sequences known to belong to this class detected by the pattern ALL, except for lentil AO.

[1] Knowles P.F., Dooley D.M. (In) Metal ions in biological systems; Sigel H., Sigel A., Eds., 30:361- 403, Marcel Dekker, New-York, (1993).

[2] Parsons M.R., Convery M.A., Wilmot C.M., Yadav K.D.S., Blakeley V., Corner A.S., Phillips S.E.V., McPherson M.J., Knowles P.F. Structure 3:1171-1184(1995).

30

103. Cys-protease (Cysteine protease)

Number of members: 358

Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences).

- 10 - Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2].
- Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2].
- Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium-activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain.
- 15 - Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3].
- Human cathepsin O [4].
- 20 - Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide).
- Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A.
- 25 - House-dust mites allergens DerP1 and EurM1.
- 30 - Cathepsin B-like proteinases from the worms *Caenorhabditis elegans* (genes *gcp-1*, *cpr-3*, *cpr-4*, *cpr-5* and *cpr-6*), *Schistosoma mansoni* (antigen SM31) and *Japonica* (antigen SJ31), *Haemonchus contortus* (genes AC-1 and AC-2), and *Ostertagia ostertagi* (CP-1 and CP-3).

- Slime mold cysteine proteinases CP1 and CP2.
- Cruzipain from *Trypanosoma cruzi* and *brucei*.
- Throphozoite cysteine proteinase (TCP) from various *Plasmodium* species.
- Proteases from *Leishmania mexicana*, *Theileria annulata* and *Theileria parva*.
- 5 - Baculoviruses cathepsin-like enzyme (v-cath).
- *Drosophila* small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain.
- Yeast thiol protease BLH1/YCP1/LAP3.
- *Caenorhabditis elegans* hypothetical protein C06G4.2, a calpain-like
- 10 protein.

Two bacterial peptidases are also part of this family:

- Aminopeptidase C from *Lactococcus lactis* (gene pepC) [5].
- 15 - Thiol protease tpr from *Porphyromonas gingivalis*.

Three other proteins are structurally related to this family, but may have lost their proteolytic activity.

- 20 - Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine.
- Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see
- 25 <PDOC00382>).
- *Plasmodium falciparum* serine-repeat protein (SERA), the major blood stage antigen. This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine.

30 The sequences around the three active site residues are well conserved and can be used as signature patterns.

152

-Consensus pattern: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]

-Consensus pattern: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH] [H is the active site residue]

5

-Consensus pattern: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue]

[1] Dufour E. Biochimie 70:1335-1342(1988).

10

[2] Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).

[3] Shi G.-P., Chapman H.A., Bhairi S.M., Deleeuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).

[4] Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).

15

[5] Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).

[6] Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).

[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

20

104. Cys_Met_Meta_PP (Cys/Met metabolism PLP-dependent enzyme)

[1] Medline: 96428687

Crystal structure of the pyridoxal-5'-phosphate dependent cystathionine beta-lyase from Escherichia coli at 1.83 Å.

25

Clausen T, Huber R, Laber B, Pohlenz HD, Messerschmidt A;
J Mol Biol 1996;262:202-224.

[1] Medline: 99059720

Crystal structure of Escherichia coli cystathionine gamma-synthase at 1.5 Å resolution.

30

Clausen T, Huber R, Prade L, Wahl MC, Messerschmidt A;
EMBO J 1998;17:6827-6838.

Database Reference: SCOP; 1cs1; fa; [SCOP-USA][CATH-PDBSUM]

This family includes enzymes involved in cysteine and

methionine metabolism. The following are members:

Cystathionine gamma-lyase,

Cystathionine gamma-synthase,

Cystathionine beta-lyase,

5 Methionine gamma-lyase,

OAH/OAS sulfhydrylase.

O-succinylhomoserine sulphhydrylase

All of these members participate in slightly different reactions.

All these enzymes use PLP (pyridoxal-5'-phosphate) as a cofactor.

10 Number of members: 52

A number of pyridoxal-dependent enzymes involved in the metabolism of cysteine, homocysteine and methionine have been shown [1,2] to be evolutionary related. These are:

- Cystathionine gamma-lyase (EC 4.4.1.1) (gamma-cystathionase), which catalyzes the transformation of cystathionine into cysteine, oxobutanoate and ammonia. This is the final reaction in the transulfuration pathway that leads from methionine to cysteine in eukaryotes.

- Cystathionine gamma-synthase (EC 4.2.99.9), which catalyzes the conversion of cysteine and succinyl-homoserine into cystathionine and succinate: the first step in the biosynthesis of methionine from cysteine in bacteria (gene metB).

- Cystathionine beta-lyase (EC 4.4.1.8) (beta-cystathionase), which catalyzes the conversion of cystathionine into homocysteine, pyruvate and ammonia: the second step in the biosynthesis of methionine from cysteine in bacteria (gene metC).

- Methionine gamma-lyase (EC 4.4.1.11) (L-methioninase) which catalyzes the transformation of methionine into methanethiol, oxobutanoate and ammonia.

- OAH/OAS sulfhydrylase, which catalyzes the conversion of acetylhomoserine into homocysteine and that of acetylserine into cysteine (gene MET17 or MET25 in yeast).

- O-succinylhomoserine sulphydrylase (EC 4.2.99.-).

- Yeast hypothetical protein YGL184c.
- Yeast hypothetical protein YHR112c.

These enzymes are proteins of about 400 amino-acid residues. The pyridoxal-P group is attached to a lysine residue located in the central section of these enzymes; the sequence around this residue is highly conserved and can be used as a signature pattern to detect this class of enzymes.

-Consensus pattern: [DQ]-[LIVMF]-x(3)-[STAGC]-[STAGC]-T-K-[FYWQ]-[LIVMF]-x-G-[HQ]-[SGNH] [K is the pyridoxal-P attachment site]

[1] Ono B.I., Tanaka K., Naito K., Heike C., Shinoda S., Yamamoto S., Ohmori S., Oshima T., Toh-E A.
J. Bacteriol. 174:3339-3347(1992).

[2] Barton A.B., Kaback D.B., Clark M.W., Keng T., Ouellette B.F.F., Storms R.K., Zeng B., Zhong W.W., Fortin N., Delaney S., Bussey H.
Yeast 9:363-369(1993).

105. Cyt_reductase

FAD/NAD-binding Cytochrome reductase

Number of members: 60

[1] Medline: 95111952

Crystal structure of the FAD-containing fragment of corn nitrate reductase at 2.5 Å resolution: relationship to other flavoprotein reductases.

Lu G, Campbell WH, Schneider G, Lindqvist Y;
Structure 1994;2:809-821.

[2] Medline: 92084635

The sequence of squash NADH:nitrate reductase and its relationship to the sequences of other flavoprotein oxidoreductases. A family of flavoprotein pyridine nucleotide cytochrome reductases.

Hyde GE, Crawford NM, Campbell WH;
J Biol Chem 1991;266:23542-23547.

- 5 106. Cytidyltrans
Phosphatidate cytidyltransferase
Number of members: 21

10 Phosphatidate cytidyltransferase (EC 2.7.7.41) [1,2,3] (also known as CDP-
diacylglycerol synthase) (CDS) is the enzyme that catalyzes the synthesis of
CDP-diacylglycerol from CTP and phosphatidate (PA). CDP-diacylglycerol is an
important branch point intermediate in both prokaryotic and eukaryotic
organisms. CDS is a membrane-bound enzyme. A conserved region located in the
C-terminal part has been selected as a signature pattern.

15 -Consensus pattern: S-x-[LIVMF]-K-R-x(4)-K-D-x-[GSA]-x(2)-[LI]-[PG]-x-H-G-G-
[LIVM]-x-D-R-[LIVMF]-D

- 20 { 1] Sparrow C.P., Raetz C.R.H.
J. Biol. Chem. 260:12084-12091(1985).
[2] Shen H., Heacock P.N., Clancey C.J., Dowhan W.
J. Biol. Chem. 271:789-795(1996).
[3] Saito S., Goto K., Tonosaki A., Kondo H.
J. Biol. Chem. 272:9503-9509(1997).

25

107. (Cytidyltransf) Cytidyltransferase. This family includes: Cholinephosphate
cytidyltransferase. Glycerol-3-phosphate cytidyltransferase.
Number of members: 64

30

- [1] Medline: 10208837 CTP:Phosphocholine Cytidyltransferase: Insights into Regulatory
Mechanisms and Novel Functions. Clement JM, Kent C; Biochem Biophys Res Commun
1999;257:643-650.

- [1] Weber I.T., Shabb J.B., Corbin J.D. *Biochemistry* 28:6122-6127(1989).
- [2] Kaupp U.B. *Trends Neurosci.* 14:150-157(1991).
- [3] Shabb J.B., Corbin J.D. *J. Biol. Chem.* 267:5723-5726(1992).

109. (cadherin)

Cadherins extracellular repeated domain signature

- 5 Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known:

- 10 - Epithelial (E-cadherin) (also known as uvomorulin or L-CAM) (CDH1).
 - Neural (N-cadherin) (CDH2).
 - Placental (P-cadherin) (CDH3).
 - Retinal (R-cadherin) (CDH4).
 - Vascular endothelial (VE-cadherin) (CDH5).
 15 - Kidney (K-cadherin) (CDH6).
 - Cadherin-8 (CDH8).
 - Osteoblast (OB-cadherin) (CDH11).
 - Brain (BR-cadherin) (CDH12).
 - T-cadherin (truncated cadherin) (CDH13).
 20 - Muscle (M-cadherin) (CDH14).
 - Liver-intestine (LI-cadherin).
 - EP-cadherin.

- 25 Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub- divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

30

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

- 5 The *Drosophila* fat protein [3] is a huge protein of over 5000 amino acids that contains 34 cadherin-like repeats in its extracellular domain.

The signature pattern that was developed for the repeated domain is located in it the C-terminal extremity which is its best conserved region. The pattern includes two conserved
10 aspartic acid residues as well as two asparagines; these residues could be implicated in the binding of calcium.

Consensus pattern[LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P Sequences known to belong to this class detected by the pattern ALL. Note this pattern is found in the first, second, and fourth
15 copies of the repeated domain. In the third copy there is a deletion of one residue after the second conserved Asp.

[1] Takeichi M. Annu. Rev. Biochem. 59:237-252(1990).

[2] Takeichi M. Trends Genet. 3:213-217(1987).

20 [3] Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J., Goodman C.S. Cell 67:853-868(1991).

110. Calreticulin family signatures

- 25 Calreticulin [1] (also known as calregulin, CRP55 or HACBP) is a high-capacity calcium-binding protein which is present in most tissues and located at the periphery of the endoplasmic (ER) and the sarcoplasmic reticulum (SR) membranes. It probably plays a role in the storage of calcium in the lumen of the ER and SR and it may well have other important functions. Structurally, calreticulin is a protein of about 400 amino acid residues consisting of
30 three domains: a) An N-terminal, probably globular, domain of about 180 amino acid residues (N-domain); b) A central domain of about 70 residues (P-domain) which contains three repeats of an acidic 17 amino acid motif. This region binds calcium with a low-capacity, but a high-affinity; c) A C-terminal domain rich in acidic residues and in lysine (C-

domain). This region binds calcium with a high-capacity but a low-affinity. Calreticulin is evolutionary related to the following proteins: - Onchocerca volvulus antigen RAL-1. RAL-1 is highly similar to calreticulin, but possesses a C-terminal domain rich in lysine and arginine and lacks acidic residues and is therefore not expected to bind calcium in that region. -

5 Calnexin [2]. A calcium-binding protein that interacts with newly synthesized glycoproteins in the endoplasmic reticulum. It seems to play a major role in the quality control apparatus of the ER by the retention of incorrectly folded proteins. - Calmegin [3] (or calnexin-T), a testis-specific calcium-binding protein highly similar to calnexin. Three signature patterns have been developed for this family of proteins. The first two patterns are based on conserved
10 regions in the N-domain; the third pattern corresponds to positions 4 to 16 of the repeated motif in the P-domain.

Consensus pattern: [KRHN]-x-[DEQN]-[DEQNK]-x(3)-C-G-G-[AG]-[FY]-[LIVM]-[KN]-[LIVMFY](2)-

Consensus pattern: [LIVM](2)-F-G-P-D-x-C-[AG]-

15 Consensus pattern: [IV]-x-D-x-[DENST]-x(2)-K-P-[DEH]-D-W-[DEN]-

[1] Michalak M., Milner R.E., Burns K., Opas M. Biochem. J. 285:681-692(1992).

[2] Bergeron J.J.M., Brenner M.B., Thomas D.Y., Williams D.B. Trends Biochem. Sci. 19:124-128(1994).

20 [3] Watanabe D., Yamada K., Nishina Y., Tajima Y., Koshimizu U., Nagata A., Nishimune Y. J. Biol. Chem. 269:7744-7749(1994).

111. Eukaryotic-type carbonic anhydrases signature (carb_anhydrase)

25 Carbonic anhydrases (EC 4.2.1.1) (CA) [1,2,3,4] are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide. Eight enzymatic and evolutionary related forms of carbonic anhydrase are currently known to exist in vertebrates: three cytosolic isozymes (CA-I, CA-II and CA-III); two membrane-bound forms (CA-IV and CA-VII); a mitochondrial form (CA-V); a secreted salivary form (CA-VI); and a yet uncharacterized isozyme [5]. In the
30 alga Chlamydomonas reinhardtii, two CA isozymes have been sequenced[6]. They are periplasmic glycoproteins evolutionary related to vertebrate CAs. Some bacteria, such as Neisseria gonorrhoeae [7] also have a eukaryotic-type CA. CAs contain a single zinc atom bound to three conserved histidine residues. As a signature for CAs, a pattern has been

developed which includes one of these zinc-binding histidines. Protein D8 from Vaccinia and other poxviruses is related to CAs but has lost two of the zinc-binding histidines as well as many otherwise conserved residues. This is also true of the N-terminal extracellular domain of some receptor-type tyrosine-protein phosphatases (see <PDOC00323>).

- 5 Consensus pattern: S-E-[HN]-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVMGA]-H-[LIVMFA](2)
[The second H is a zinc ligand]-

Note: most prokaryotic CA's as well as plant chloroplast CA's belong to another, evolutionary distinct family of proteins (see <PDOC00586

- 10 [1] Deutsch H.F. Int. J. Biochem. 19:101-113(1987).
[2] Fernley R.T. Trends Biochem. Sci. 13:356-359(1988).
[3] Tashian R.E. BioEssays 10:186-192(1989).
[4] Edwards Y. Biochem. Soc. Trans. 18:171-175(1990).
[5] Skaggs L.A., Bergenheim N.C.H., Venta P.J., Tashian R.E. Gene 126:291-292(1993).
15 [6] Fujiwara S., Fukuzawa H., Tachiki A., Miyachi S. Proc. Natl. Acad. Sci. U.S.A. 87:9779-9783(1990).
[7] Huang S., Xue Y., Sauer-Eriksson E., Chirica L., Lindskog S., Jonsson B.H. 23.CO:2-"J. Mol. Biol. 283:301-310(1998).

20 112. Caseins alpha/beta signature

Caseins [1] are the major protein constituent of milk. Caseins can be classified into two families; the first consists of the kappa-caseins, and the second groups the alpha-s1, alpha-s2, and beta-caseins. The alpha/beta caseins are a rapidly diverging family of proteins. However

25 two regions are conserved: a cluster of phosphorylated serine residues and the signal sequence. The signature pattern has been developed for this family of proteins based upon the last eight residues of the signal sequence.

Consensus pattern: C-L-[LV]-A-x-A-[LVF]-A -

- 30 [1] Holt C., Sawyer L. Protein Eng. 2:251-259(1988).

113. Catalase signatures

Consensus pattern: [IF]-x-[RH]-x(4)-[EQ]-R-x(2)-H-x(2)-[GAS]-[GASTF]-[GAST] [H is an active site residue]

[1] Murthy M.R.N., Reid T.J. III, Sicignano A., Tanaka N., Rossmann M.G. J. Mol. Biol. 152:465-499(1981).

[2] Melik-Adamyany W.R., Barynin V.V., Vagin A.A., Borisov V.V., Vainshtein B.K., Fita I., Murthy M.R.N., Rossmann M.G. J. Mol. Biol. 188:63-72(1986).

[3] von Ossowski I., Hausner G., Loewen P.C. J. Mol. Evol. 37:71-76(1993).

A conserved domain of 43 amino acids is found in several plant and fungal proteins that have a common binding specificity for oligosaccharides of N-acetylglucosamine [1]. This domain may be involved in the recognition or binding of chitin subunits. It has been found in the proteins listed below. - A number of non-leguminous plant lectins. The best characterized of these lectins are the three highly homologous wheat germ agglutinins (WGA-1, 2 and 3). WGA is an N-acetylglucosamine/N-acetylneuraminic acid binding lectin which structurally consists of a fourfold repetition of the 43 amino acid domain. The same type of structure is found in a barley root-specific lectin as well as a rice lectin. - Plants endochitinases (EC 3.2.1.14) from class IA (see <PDOC00620>). Endochitinases are enzymes that catalyze the hydrolysis of the beta-1,4 linkages of N-acetyl glucosamine polymers of chitin. Plant

chitinases function as a defense against chitin containing fungal pathogens. Class IA chitinases generally contain one copy of the chitin-binding domain at their N-terminal extremity. An exception is agglutinin/chitinase [2] from the stinging nettle *Urtica dioica* which contains two copies of the domain. - Hevein [5], a wound-induced protein found in the latex of rubber trees. - Win1 and win2, two wound-induced proteins from potato. - Kluyveromyces lactis killer toxin alpha subunit [3]. The toxin encoded by the linear plasmid pGKL1 is composed of three subunits: alpha, beta, and gamma. The gamma subunit harbors toxin activity and inhibits growth of sensitive yeast strains in the G1 phase of the cell cycle; the alpha subunit, which is proteolytically processed from a larger precursor that also contains the beta subunit, is a chitinase (see <PDOC00839>). In chitinases, as well as in the potato wound-induced proteins, the 43-residue domain directly follows the signal sequence and is therefore at the N-terminal of the mature protein; in the killer toxin alpha subunit it is located in the central section of the protein. The domain contains eight conserved cysteine residues which have all been shown, in WGA, to be involved in disulfide bonds. The topological arrangement of the four disulfide bonds is shown in the following figure: +-----
 -----+ +---|-----+ |||| xx CgxxxxxxxxCxxxxCCsxxgxCgxxxxxCxxx CxxxxC |
 *****|***** |||| +-----+ 'C': conserved cysteine involved in a
 disulfide bond. '*': position of the pattern.

-Consensus pattern: C-x(4,5)-C-C-S-x(2)-G-x-C-G-x(4)-[FYW]-C [The five C's are involved in disulfide bonds]

[1] Wright H.T., Sandrasegaram G., Wright C.S. J. Mol. Evol. 33:283-294(1991).

[2] Lerner D.R., Raikhel N.V. J. Biol. Chem. 267:11085-11091(1992).

[3] Butler A.R., O'Donnel R.W., Martin V.J., Gooday G.W., Stark M.J.R. Eur. J. Biochem. 199:483-488(1991).

115. (Chitinase 1) Chitinases family 19 signatures

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the viewpoint of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2, E1]. Chitinases of family 19 (also known as classes IA or I and IB or II) are enzymes from plants

that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (see the relevant entry <PDOC00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues. Two highly conserved regions have been selected as signature patterns, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond.

Consensus pattern: C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F- [GSA]

Consensus pattern: [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM]

[1] Flach J., Pilet P.-E., Jolles P. *Experientia* 48:701-716(1992).

[2] Henrissat B. *Biochem. J.* 280:309-316(1991).

116. chloroa_b-bind

Chlorophyll A-B binding proteins. Number of members: 211

117. chromo

The 'chromo' (CHRromatin Organization MOdifier) domain [1 to 4] is a conserved region of about 60 amino acids which was originally found in *Drosophila* modifiers of variegation, which are proteins that modify the structure of chromatin to the condensed morphology of heterochromatin, a cytologically visible condition where gene expression is repressed. In protein Polycomb, the chromo domain has been shown to be important for chromatin targeting. Proteins that contains a chromo domain seem to fall into three classes:

- a) Proteins which have a N-terminal chromo domain followed by a region which is related to but distinct from the chromo domain and which has been termed [3] the 'chromo shadow' domain.
- b) Proteins with a single chromo domain.
- c) Proteins with paired tandem chromo domains.

Currently, this domain has been found in the following proteins:

Class A.

- 5 - Drosophila heterochromatin protein Su(var)205 (HP1).
- Human heterochromatin protein HP1 alpha.
- Mammalian modifier 1 and modifier 2.
- Fission yeast swi6, a protein involved in the repression of the silent mating-type loci mat2 and mat3.

Class B.

- Drosophila protein Polycomb (Pc).
- Mammalian modifier 3, a homolog of Pc.
- Drosophila protein Su(var)3-9, a suppressor of position-effect variegation.
- 15 - Human Mi-2 autoantigen, characteristic of dermatomyositis.
- Fungal retrotransposon polyproteins: 'skippy' from *Fusarium oxysporum*, 'grasshopper' and 'MAGGY' from *Magnaporthe grisea* and Cft-1 from *Cladosporium fulvum*.
- Fission yeast hypothetical protein SpAC18G6.02c.
- 20 - *Caenorhabditis elegans* hypothetical protein C29H12.5
- *Caenorhabditis elegans* hypothetical protein ZK1236.2.
- *Caenorhabditis elegans* hypothetical protein T09A5.8.

Class C.

- 25 - Mammalian DNA-binding/helicase proteins CHD-1 to CHD-4.
- Yeast protein CHD1.

The signature pattern for this domain corresponds to its best conserved section, which is located in its central part.

30 - Consensus pattern: [FYL]-x-[LIVMC]-[KR]-W-x-[GDNR]-[FYWLME]-x(5,6)-[ST]-W-[ESV]-[PSTDEN]-x(2,3)-[LIVMC]

[1] Paro R. Trends Genet. 6:416-421(1990).

[2] Singh P.B., Miller J.R., Pearce J., Kothary R., Burton R.D., Paro R., James T.C., Gaunt S.J. Nucleic Acids Res. 19:789-794(1991).

[3] Aasland R., Stewart A.F. Nucleic Acids Res. 23:3168-3173(1995).

5 [4] Koonin E.V., Zhou S., Lucchesi J.C. Nucleic Acids Res. 23:4229-4233(1995).

118. citrate_synth

10 Citrate synthase (EC 4.1.3.7) (CS) is the tricarboxylic acid cycle enzyme that catalyzes the synthesis of citrate from oxaloacetate and acetyl-CoA in an aldol condensation. CS can directly form a carbon-carbon bond in the absence of metal ion cofactors.

15 In prokaryotes, citrate synthase is composed of six identical subunits. In eukaryotes, there are two isozymes of citrate synthase: one is found in the mitochondrial matrix, the second is cytoplasmic. Both seem to be dimers of identical chains.

20 There are a number of regions of sequence similarity between prokaryotic and eukaryotic citrate synthases. One of the best conserved contains a histidine which is one of three residues shown [1] to be involved in the catalytic mechanism of the vertebrate mitochondrial enzyme. This region has been used as a signature pattern.

25 -Consensus pattern: G-[FYA]-[GA]-H-x-[IV]-x(1,2)-[RKT]-x(2)-D-[PS]-R [H is an active site residue]

[1] Karpusas M., Branchaud B., Remington S.J. Biochemistry 29:2213-2219(1990).

30

119. clpA_B

Chaperonin clpA/B

CAUTION! This family is a subfamily of the AAA

superfamily. The threshold has been set very high to stop overlaps with the AAA superfamily. This entry will be subsumed by AAA in the future.

Number of members: 39

5

A number of ATP-binding proteins that are thought to protect cells from extreme stress by controlling the aggregation or denaturation of vital cellular structures have been shown [1,2] to be evolutionary related. These proteins are listed below.

10

- *Escherichia coli* clpA, which acts as the regulatory subunit of the ATP-dependent protease clp.

- *Rhodopseudomonas blastica* clpA homolog.

- *Escherichia coli* heat shock protein clpB and homologs in other bacteria.

15

- *Bacillus subtilis* protein mecB.

- Yeast heat shock protein 104 (gene HSP104), which is vital for tolerance to heat, ethanol and other stresses.

- *Neurospora* heat shock protein hsp98.

- Yeast mitochondrial heat shock protein 78 (gene HSP78) [3].

20

- CD4A and CD4b, two highly related tomato proteins that seem to be located in the chloroplast.

- *Trypanosoma brucei* protein clp.

- *Porphyra purpurea* chloroplast encoded clpC.

25

The size of these proteins range from 84 Kd (clpA) to slightly more than 100 Kd (HSP104). They all share two conserved regions of about 200 amino acids that each contains an ATP-binding site. In addition to the ATP-binding A and B motifs there are many parts in these two domains that are also conserved. Two of these regions have been selected as signature patterns. The first signature is located in the first domain, some ten residues to the C-terminal of the ATP-binding B motif. The second pattern is located in the second domain in-between the ATP-binding A and B motifs.

30

167

-Consensus pattern: D-[AI]-[SGA]-N-[LIVMF](2)-K-[PT]-x-L-x(2)-G

-Consensus pattern: R-[LIVMFY]-D-x-S-E-[LIVMFY]-x-E-[KRQ]-x-[STA]-x-[STA]-[KR]-[LIVM]-x-G-[STA]

- 5 [1] Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M., Mattick J.S., Dalrymple B., Kuramitsu H., Shiroza T., Foster T., Clark W.P., Ross B., Squires C.L., Maurizi M.R. Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).
- [2] Parsell D.A., Sanchez Y., Stitzel J.D., Lindquist S. Nature 353:270-273(1991).
- [3] Leonhardt S.A., Fearon K., Danese P.N., Mason T.L. Mol. Cell. Biol. 13:6304-6313(1993).

120. cofilin_ADF

Cofilin/tropomyosin-type actin-binding proteins

[1]

Medline: 97290449

Structure determination of yeast cofilin.

Fedorov AA, Lappalainen P, Fedorov EV, Drubin DG, Almo SC;

Nat Struct Biol 1997;4:366-369.

[2]

Medline: 97290450

Crystal structure of the actin-binding protein actophorin from Acanthamoeba.

Leonard SA, Gittis AG, Petrella EC, Pollard TD, Lattman EE;

Nat Struct Biol 1997;4:369-373.

[3]

Medline: 97420794

F-actin and G-actin binding are uncoupled by mutation of conserved tyrosine residues in maize actin depolymerizing factor.

Jiang CJ, Weeds AG, Khan S, Hussey PJ;

Proc Natl Acad Sci U S A 1997;94:9973-9978.

[4]

Medline: 97357155

Cofilin promotes rapid actin filament turnover in vivo.

Lappalainen P, Drubin DG;

Nature 1997;388:78-82.

5 Severs actin filaments and binds to actin monomers.

Number of members: 44

Actin-depolymerizing proteins sever actin filaments (F-actin) and/or bind to actin monomers, or G-actin, thus preventing actin-polymerization by sequestering the monomers. The following proteins are evolutionary related and belong to a family of low molecular weight (137 to 166 residues) actin-depolymerizing proteins [1,2,3,4]:

- Cofilin from vertebrates, slime mold and yeast. Cofilin binds to F-actin and acts as a pH-dependent actin-depolymerizing protein.
- Destrin from vertebrates. Destrin binds to G-actin in a pH-independent manner and prevents polymerization.
- *Caenorhabditis elegans* unc-60.
- *Acanthamoeba castellanii* actophorin.
- Plants actin depolymerizing factor (ADF).

The most conserved region of these proteins is a twenty amino-acid segment that ends some 30 residues from their C-terminal extremity. This segment has been shown [5] to be important for actin-binding.

-Consensus pattern: P-[DE]-x-[SA]-x-[LIVMT]-[KR]-x-[KR]-M-[LIVM]-[YA]-[STA](3)-x(3)-[LIVMF]-[KR]

[1] Hawkins M., Pope B., MacIver S.K., Weeds A.G. Biochemistry 32:9985-9993(1993).

[2] Iida K., Moriyama K., Matsumoto S., Kawasaki H., Nishida E., Yahara I. Gene 124:115-120(1993).

[3] Quirk S., MacIver S.K., Ampe C., Doberstein S.K., Kaiser D.A., van Damme J., Vandekerckhove J., Pollard T.D. Biochemistry 32:8525-8533(1993).

[4] McKim K.S., Matheson C., Marra M.A., Wakarchuk M.F., Baillie D.L. Mol. Gen. Genet. 242:346-357(1994).

[5] Moriyama K., Yonezawa N., Sakai H., Yahara I., Nishida E. J. Biol. Chem. 267:7240-7244(1992).

5

121. (Complex 24kd) Respiratory-chain NADH dehydrogenase 24 Kd subunit signature
Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or
NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner
mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as
a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this
bioenergetic enzyme complex there is one with a molecular weight of 24 Kd (in mammals),
which is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 2Fe-
2S iron-sulfur cluster. The 24 Kd subunit is nuclear encoded, as a precursor form with a
transit peptide in mammals, and in *Neurospora crassa*. The 24 Kd subunit is highly similar to
[3,4]: - Subunit E of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoE*). -
Subunit NQO2 of *Paracoccus denitrificans* NADH-ubiquinone oxidoreductase. A highly
conserved region, located in the central section of this subunit containing two conserved
cysteines that are probably involved in the binding of the 2Fe-2S center has been selected as a
signature pattern.

-Consensus pattern: D-x(2)-F-[ST]-x(5)-C-L-G-x-C-x(2) [GA]-P [The two C's are putative
2Fe-2S ligands]

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

[3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).

[4] Weidner U., Geier S., Ptocek A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

30

122. copper-bind

Copper binding proteins, plastocyanin/azurin family

Number of members: 70

Blue or 'type-1' copper proteins are small proteins which bind a single copper atom and which are characterized by an intense electronic absorption band near 600 nm [1,2]. The most well known members of this class of proteins are the plant chloroplastic plastocyanins, which exchange electrons with cytochrome c6, and the distantly related bacterial azurins, which exchange electrons with cytochrome c551. This family of proteins also includes all the proteins listed below (references are only provided for recently determined sequences).

- Amicyanin from bacteria such as *Methylobacterium extorquens* or *Thiobacillus versutus* that can grow on methylamine. Amicyanin appears to be an electron receptor for methylamine dehydrogenase.
- Auracyanins A and B from *Chloroflexus aurantiacus* [3]. These proteins can donate electrons to cytochrome c-554.
- Blue copper protein from *Alcaligenes faecalis*.
- Cupredoxin (CPC) from cucumber peelings [4].
- Cusacyanin (basic blue protein; plantacyanin, CBP) from cucumber.
- Halocyanin from *Natrobacterium pharaonis* [5], a membrane associated copper-binding protein.
- Pseudoazurin from *Pseudomonas*.
- Rusticyanin from *Thiobacillus ferrooxidans*. Rusticyanin is an electron carrier from cytochrome c-552 to the α -type oxidase [6].
- Stellacyanin from the Japanese lacquer tree.
- Umecyanin from horseradish roots.
- Allergen Ra3 from ragweed. This pollen protein is evolutionary related to the above proteins, but seems to have lost the ability to bind copper.

Although there is an appreciable amount of divergence in the sequence of all these proteins, the copper ligand sites are conserved and a pattern which includes two of the ligands (a cysteine and a histidine) has been developed.

-Consensus pattern: [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ] [C and H are copper ligands]

- [1] Garret T.P.J., Clingeffer D.J., Guss J.M., Rogers S.J., Freeman H.C. J. Biol. Chem. 259:2822-2825(1984).
- [2] Ryden L.G., Hunt L.T. J. Mol. Evol. 36:41-66(1993).
- [3] McManus J.D., Brune D.C., Han J., Sanders-Loehr J., Meyer T.E., Cusanovich M.A., Tollin G., Blankenship R.E. J. Biol. Chem. 267:6531-6540(1992).
- [4] Mann K., Schaefer W., Thoenes U., Messerschmidt A., Mehrabian Z., Nalbandyan R. FEBS Lett. 314:220-223(1992).
- [5] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterheld D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).
- [6] Yano T., Fukumori Y., Yamanaka T. FEBS Lett. 288:159-162(1991).

123. Chaperonins cpn10 signature

Chaperonins [1,2] are proteins involved in the folding of proteins or the assembly of oligomeric protein complexes. They seem to assist other polypeptides in maintaining or assuming conformations which permit their correct assembly into oligomeric structures. They are found in abundance in prokaryotes, chloroplasts and mitochondria. Chaperonins form oligomeric complexes and are composed of two different types of subunits: a 60 Kd protein, known as cpn60 (groEL in bacteria) and a 10 Kd protein, known as cpn10 (groES in bacteria). The cpn10 protein binds to cpn60 in the presence of MgATP and suppresses the ATPase activity of the latter. Cpn10 is a protein of about 100 amino acid residues whose sequence is well conserved in bacteria, vertebrate mitochondria and plants chloroplast [3,4]. Cpn10 assembles as an heptamer that forms a dome [5]. As a signature pattern for cpn10, a region located in the N-terminal section of the protein was selected.

Consensus pattern: [LIVMFY]-x-P-[ILT]-x-[DEN]-[KR]-[LIVMFA](3)-[KREQ]-x(8,9)-[SG]-x-[LIVMFY](3)-

Note: this pattern is found twice in the plant chloroplast protein which consist of the tandem repeat of a cpn10 domain

- [1] Ellis R.J., van der Vies S.M. Annu. Rev. Biochem. 60:321-347(1991).
 [2] Zeilsta-Ryalls J., Fayet O., Georgopoulos C. Annu. Rev. Microbiol. 45:301-325(1991).
 [3] Hartman D.J., Hoogenraad N.J., Condrón R., Hoj P.B. Proc. Natl. Acad. Sci. U.S.A. 89:3394-3398(1992).
 5 [4] Bertsch U., Soll J., Seetharam R., Viitanen P.V. Proc. Natl. Acad. Sci. U.S.A. 89:8696-8700(1992).
 [5] Hunt J.F., Weaver A.J., Landry S.J., Gierasch L., Deisenhofer J. Nature 379:37-45(1996).

124. Chaperonins cpn60 signature (cpn60_TCP1)

Chaperonins [1,2] are proteins involved in the folding of proteins or the assembly of oligomeric protein complexes. Their role seems to be to assist other polypeptides to maintain or assume conformations which permit their correct assembly into oligomeric structures. They are found in abundance in prokaryotes, chloroplasts and mitochondria. Chaperonins form oligomeric complexes and are composed of two different types of subunits: a 60 Kd protein, known as cpn60 (groEL in bacteria) and a 10 Kd protein, known as cpn10 (groES in bacteria). The cpn60 protein shows weak ATPase activity and is a highly conserved protein of about 550 to 580 amino acid residues which has been described by different names in different species: - Escherichia coli groEL protein, which is essential for the growth of the bacteria and the assembly of several bacteriophages. - Cyanobacterial groEL analogues. - Mycobacterium tuberculosis and leprae 65 Kd antigen, Coxiella burnetti heat shock protein B (gene htpB), Rickettsia tsutsugamushi major antigen 58, and Chlamydial 57 Kd hypersensitivity antigen (gene hypB). - Chloroplast RuBisCO subunit binding-protein alpha and beta chains, which bind ribulose biphosphate carboxylase small and large subunits and are implicated in the assembly of the enzyme oligomer. - Mammalian mitochondrial matrix protein P1 (mitonin or P60). - Yeast HSP60 protein, a mitochondrial assembly factor. As a signature pattern for these proteins, a rather well-conserved region of twelve residues, located in the last third of the cpn60sequence was chosen.

Consensus pattern: A-[AS]-x-[DEQ]-E-x(4)-G-G-[GA]-

- [1] Ellis R.J., van der Vies S.M. Annu. Rev. Biochem. 60:321-347(1991).
 [2] Zeilsta-Ryalls J., Fayet O., Georgopoulos C. Annu. Rev. Microbiol. 45:301-325(1991).

Chaperonins TCP-1 signatures (cpn60_TCP1)

The TCP-1 protein [1,2] (Tailless Complex Polypeptide 1) was first identified in mice where it is especially abundant in testis but present in all cell types. It has since been found and characterized in many other mammalian species, in *Drosophila* and in yeast. TCP-1 is a highly conserved protein of about 60 Kd (556 to 560 residues) which participates in a hetero-oligomeric 900 Kd double-torus shaped particle [3] with 6 to 8 other different subunits. These subunits, the chaperonin containing TCP-1 (CCT) subunit beta, gamma, delta, epsilon, zeta and eta are evolutionary related to TCP-1 itself [4,5]. The CCT is known to act as a molecular chaperone for tubulin, actin and probably some other proteins. The CCT subunits are highly related to archeobacterial counterparts: - TF55 and TF56 [6], a molecular chaperone from *Sulfolobus shibatae*. TF55 has ATPase activity, is known to bind unfolded polypeptides and forms a oligomeric complex of two stacked nine-membered rings. - Thermosome [7], from *Thermoplasma acidophilum*. The thermosome is composed of two subunits (alpha and beta) and also seems to be a chaperone with ATPase activity. It forms an oligomeric complex of eight-membered rings. The TCP-1 family of proteins are weakly, but significantly [8], related to the cpn60/groEL chaperonin family (see <PDOC00268>). As signature patterns of this family of chaperonins, three conserved regions located in the N-terminal domain were chosen.

Consensus pattern: [RKEL]-[ST]-x-[LMFY]-G-P-x-[GSA]-x-x-K-[LIVMF](2)-

Consensus pattern: [LIVM]-[TS]-[NK]-D-[GA]-[AVNHNK]-[TAV]-[LIVM](2)-x(2)-[LIVM]-x-[LIVM]-x-[SNH]-[POH]-

Consensus pattern: Q-[DEK]-x-x-[LIVMGTA]-[GA]-D-G-T-

[1] Ellis J. Nature 358:191-192(1992).

[2] Nelson R.J., Craig E.A. Curr. Biol. 2:487-489(1992).

[3] Lewis V.A., Hynes G.M., Zheng D., Saibil H., Willison K.R. Nature 358:249-252(1992).

[4] Kubota H., Hynes G., Carne A., Ashworth A., Willison K.R. Curr. Biol. 4:89-99(1994)

[5] Kim S., Willison K.R., Horwich A.L. Trends Biochem. Sci. 20:543-548(1994).

[6] Trent J.D., Nimmesgern E., Wall J.S., Hartl F.U., Horwich A.L. Nature 354:490-493(1991).

- [7] Waldmann T., Lupas A., Kellermann J., Peters J., Baumeister W. Biol. Chem. Hoppe-Seyler 376:119-126(1995).
- [8] Hemmingsen S.M. Nature 357:650-650(1992).

[8] Hemmingsen S.M. Nature 357:650-650(1992).

5

125. cyclin (Cyclins)

The cyclins include an internal duplication, which is related to that found in TFIIB and the RB protein.

[1]

10 Medline: 94203808

Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107.

Gibson TJ, Thompson JD, Blocker A, Kouzarides T;
Nucleic Acids Res 1994;22:946-952.

15

[2]

Medline: 96164440

The crystal structure of cyclin A

Brown NR, Noble MEM, Endicott JA, Garman EF, Wakatsuki S,
Mitchell E, Rasmussen B, Hunt T, Johnson LN;

20

Structure. 1995;3:1235-1247.

Complex of cyclin and cyclin dependant kinase.

[3]

Medline: 96313126

Structural basis of cyclin-dependant kinase activation by phosphorylation.

25

Russo AA, Jeffrey PD, Pavletich NP;

Nat Struct Biol. 1996;3:696-700.

Cyclins regulate cyclin dependant kinases (CDKs).

The most divergent prosite members have been included. Swiss:P22674 the Uracil-DNA glycosylase 2 is the highest noise and may be related but has not been included.

30

the Uracil-DNA glycosylase 2 is the highest noise and may be related but has not been included.

Number of members: 189

Cyclins [1,2,3] are eukaryotic proteins which play an active role in controlling nuclear cell division cycles. Cyclins, together with the p34 (cdc2) or cdk2 kinases, form the Maturation Promoting Factor (MPF). There are two main groups of cyclins:

- G2/M cyclins, essential for the control of the cell cycle at the G2/M (mitosis) transition. G2/M cyclins accumulate steadily during G2 and are abruptly destroyed as cells exit from mitosis (at the end of the M-phase).
- G1/S cyclins, essential for the control of the cell cycle at the G1/S (start) transition.

In most species, there are multiple forms of G1 and G2 cyclins. For example, in vertebrates, there are two G2 cyclins, A and B, and at least three G1 cyclins, C, D, and E.

A cyclin homolog has also been found in herpesvirus saimiri [4].

The best conserved region is in the central part of the cyclins' sequences, known as the 'cyclin-box'. From this, a 32 residue pattern has been derived.

-Consensus pattern: R-x(2)-[LIVMSA]-x(2)-[FYWS]-[LIVM]-x(8)-[LIVMFC]-x(4)-[LIVMFYA]-x(2)-[STAGC]-[LIVMFYQ]-x-[LIVMFYC]-[LIVMFY]-D-[RKH]-[LIVMFYW]

[1] Nurse P. Nature 344:503-508(1990).

[2] Norbury C., Nurse P. Curr. Biol. 1:23-24(1991).

[3] Lew D.J., Reed S.I. Trends Cell Biol. 2:77-81(1992).

[4] Nicholas J., Cameron K.R., Honess R.W. Nature 355:362-365(1992).

126. Cystatin domain

This is a very diverse family. Attempts to define separate subfamilies have failed. Typically, either the N-terminal or C-terminal end is very divergent. But splitting into two domains

would make very short families. Cathelicidins are related to this family but have not been included. Number of members: 147

Inhibitors of cysteine proteases [1,2,3], which are found in the tissues and body fluids of animals, in the larva of the worm *Onchocerca volvulus* [4], as well as in plants, can be grouped into three distinct but related families:

- Type 1 cystatins (or stefins), molecules of about 100 amino acid residues with neither disulfide bonds nor carbohydrate groups.
- Type 2 cystatins, molecules of about 115 amino acid residues which contain one or two disulfide loops near their C-terminus.
- Kininogens, which are multifunctional plasma glycoproteins.

They are the precursor of the active peptide bradykinin and play a role in blood coagulation by helping to position optimally prekallikrein and factor XI next to factor XII. They are also inhibitors of cysteine proteases. Structurally, kininogens are made of three contiguous type-2 cystatin domains, followed by an additional domain (of variable length) which contains the sequence of bradykinin. The first of the three cystatin domains seems to have lost its inhibitory activity.

In all these inhibitors, there is a conserved region of five residues which has been proposed to be important for the binding to the cysteine proteases. The consensus pattern starts one residue before this conserved region.

-Consensus pattern: [GSTEQKRV]-Q-[LIVT]-[VAF]-[SAGQ]-G-x-[LIVMNK]-x(2)-[LIVMFY]-x-[LIVMFYA]-[DENOKRHSIV]

[1] Barrett A.J. Trends Biochem. Sci. 12:193-196(1987).

[2] Rawlings N.D., Barrett A.J. J. Mol. Evol. 30:60-71(1990).

[3] Turk V., Bode W. FEBS Lett. 285:213-219(1991).

[4] Lustigman S., Brotman B., Huima T., Prince A.M. Mol. Biochem. Parasitol. 45:65-76(1991).

127. cytochrome_c (Cytochrome c)

The Pfam entry does not include all prosite members.

The cytochrome 556 and cytochrome c' families are

not included.

Number of members: 259

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

-Consensus pattern: C-{CPWHF}-{CPWR}-C-H-{CFYW}

[1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

128. (DAGKa) Diacylglycerol kinase accessory domain (presumed)

Diacylglycerol (DAG) is a second messenger that acts as a protein kinase C activator. This domain is assumed to be an accessory domain: its function is unknown.

[1] Sakane F, Yamada K, Kanoh H, Yokoyama C, Tanabe T, Nature 1990;344:345-348.[2] Sakane F, Imai S, Kai M, Wada I, Kanoh H, J Biol Chem 1996;271:8394-8401.[3] Schaap D, de Widt J, van der Wal J, Vandekerckhove J, van, Damme J, Gussow D, Ploegh HL, van Blitterswijk WJ, van der, Bend RL, FEBS Lett 1990;275:151-158. [4] Kanoh H, Yamada K, Sakane F, Trends Biochem Sci 1990;15:47-50.

129. (DAGKc) Diacylglycerol kinase catalytic domain (presumed)

Diacylglycerol (DAG) is a second messenger that acts as a protein kinase C activator. The catalytic domain is assumed from the finding of bacterial homologues.

[1] Sakane F, Yamada K, Kanoh H, Yokoyama C, Tanabe T, Nature 1990;344:345-348. [2] Sakane F, Imai S, Kai M, Wada I, Kanoh H, J Biol Chem 1996;271:8394-8401. [3] Schaap D, de Widt J, van der Wal J, Vandekerckhove J, van, Damme J, Gussow D, Ploegh

HL, van Blitterswijk WJ, van der, Bend RL, FEBS Lett 1990;275:151-158. [4] Kanoh H, Yamada K, Sakane F, Trends Biochem Sci 1990;15:47-50.

130. D-amino acid oxidases signature(DAO)

D-amino acid oxidase (EC 1.4.3.3) (DAMOX or DAO) is an FAD flavoenzyme that catalyzes the oxidation of neutral and basic D-amino acids into their corresponding keto acids. DAOs have been characterized and sequenced in fungi and vertebrates where they are known to be located in the peroxisomes. D-aspartate oxidase (EC 1.4.3.1) (DASOX) [1] is an enzyme, structurally related to DAO, which catalyzes the same reaction but is active only toward dicarboxylic D-amino acids. In DAO, a conserved histidine has been shown [2] to be important for the enzyme's catalytic activity. The conserved region around this residue has been developed as a signature pattern for these enzymes.

Consensus pattern: [LIVM](2)-H-[NHA]-Y-G-x-[GSA](2)-x-G-x(5)-G-x-A [H is a probable active site residue]o-

[1] Negri A., Ceciliani F., Tedeschi G., Simonc T., Ronchi S. J. Biol. Chem. 267:11865-11871(1992).

[2] Miyano M., Fukui K., Watanabe F., Takahashi S., Tada M., Kanashiro M., Miyake Y. J. Biochem. 109:171-177(1991).

131. DEAD and DEAH box families ATP-dependent helicases signatures

A number of eukaryotic and prokaryotic proteins have been characterized [1,2,3] on the basis of their structural similarity. They all seem to be involved in ATP-dependent, nucleic-acid unwinding. Proteins currently known to belong to this family are: - Initiation factor eIF-4A. Found in eukaryotes, this protein is a subunit of a high molecular weight complex involved in 5'cap recognition and the binding of mRNA to ribosomes. It is an ATP-dependent RNA-helicase. - PRP5 and PRP28. These yeast proteins are involved in various ATP-requiring steps of the pre-mRNA splicing process. - P110, a mouse protein expressed specifically during spermatogenesis. - An3, a Xenopus putative RNA helicase, closely related to P110. - SPP81/DED1 and DBP1, two yeast proteins probably involved in pre-mRNA splicing and

related to P110. - *Caenorhabditis elegans* helicase *glh-1*. - MSS116, a yeast protein required for mitochondrial splicing. - SPB4, a yeast protein involved in the maturation of 25S ribosomal RNA. - p68, a human nuclear antigen. p68 has ATPase and DNA-helicase activities in vitro. It is involved in cell growth and division. - Rm62 (p62), a *Drosophila* putative RNA helicase related to p68. - DBP2, a yeast protein related to p68. - DHH1, a yeast protein. - DRS1, a yeast protein involved in ribosome assembly. - MAK5, a yeast protein involved in maintenance of dsRNA killer plasmid. - ROK1, a yeast protein. - *ste13*, a fission yeast protein. - Vasa, a *Drosophila* protein important for oocyte formation and specification of embryonic posterior structures. - Mc31B, a *Drosophila* maternally expressed protein of unknown function. - *dbpA*, an *Escherichia coli* putative RNA helicase. - *deaD*, an *Escherichia coli* putative RNA helicase which can suppress a mutation in the *rpsB* gene for ribosomal protein S2. - *rhlB*, an *Escherichia coli* putative RNA helicase. - *rhlE*, an *Escherichia coli* putative RNA helicase. - *srmB*, an *Escherichia coli* protein that shows RNA-dependent ATPase activity. It probably interacts with 23S ribosomal RNA. - *Caenorhabditis elegans* hypothetical proteins T26G10.1, ZK512.2 and ZK686.2. - Yeast hypothetical protein YHR065c. - Yeast hypothetical protein YHR169w. - Fission yeast hypothetical protein SpAC31A2.07c. - *Bacillus subtilis* hypothetical protein *yxjN*. All these proteins share a number of conserved sequence motifs. Some of them are specific to this family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' [4,E1]. One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins [3,5,6,E1]. Proteins currently known to belong to this subfamily are: - PRP2, PRP16, PRP22 and PRP43. These yeast proteins are all involved in various ATP-requiring steps of the pre-mRNA splicing process. - Fission yeast *prh1*, which may be involved in pre-mRNA splicing. - Male-less (*mle*), a *Drosophila* protein required in males, for dosage compensation of X chromosome linked genes. - RAD3 from yeast. RAD3 is a DNA helicase involved in excision repair of DNA damaged by UV light, bulky adducts or cross-linking agents. Fission yeast *rad15* (*rhp3*) and mammalian DNA excision repair protein XPD (ERCC-2) are the homologs of RAD3. - Yeast CHL1 (or CTF1), which is important for chromosome transmission and normal cell cycle progression in G(2)/M. - Yeast TPS1. - Yeast hypothetical protein YKL078w. - *Caenorhabditis elegans* hypothetical proteins C06E1.10 and K03H1.2. - Poxviruses' early transcription factor 70 Kd subunit which acts with RNA polymerase to

initiate transcription from early gene promoters. - 18, a putative vaccinia virus helicase. - hrpA, an Escherichia coli putative RNA helicase. Signature patterns for both subfamilies were developed.

5 Consensus pattern: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN

Consensus pattern: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR]

Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see the relevant entry <[PDOC00017](#)

10 [1] Schmid S.R., Linder P. Mol. Microbiol. 6:283-292(1992).

[2] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[3] Wassarman D.A., Steitz J.A. Nature 349:463-464(1991).

[4] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

15 [5] Harosh I., Deschavanne P. Nucleic Acids Res. 19:6331-6331(1991).

[6] Koonin E.V., Senkevich T.G. J. Gen. Virol. 73:989-993(1992).

132. (DHP synthase) 3,4-dihydroxy-2-butanone 4-phosphate synthase

20 3,4-Dihydroxy-2-butanone 4-phosphate is biosynthesized from ribulose 5-phosphate and serves as the biosynthetic precursor for the xylene ring of riboflavin. Sometimes found as a bifunctional enzyme with [GTP_cyclohydro2](#).

Richter G, Krieger C, Volk R, Kis K, Ritz H, Gotze E, Bacher A, Methods Enzymol 1997;280:374-382.

25

133. (DHDPS) Dihydrodipicolinate synthetase signatures

Dihydrodipicolinate synthetase (EC [4.2.1.52](#)) (DHDPS) [1] catalyzes, in higher plants chloroplast and in many bacteria (gene dapA), the first reaction specific to the biosynthesis of lysine and of diaminopimelate. DHDPS is responsible for the condensation of aspartate
30 semialdehyde and pyruvate by aping-pong mechanism in which pyruvate first binds to the enzyme by forming a Schiff-base with a lysine residue. Three other proteins are structurally related to DHDPS and probably also act via a similar catalytic mechanism: - Escherichia coli

N-acetylneuraminate lyase (EC 4.1.3.3) (gene nanA), which catalyzes the condensation of N-acetyl-D-mannosamine and pyruvate to form N-acetylneuraminate. - Rhizobium meliloti protein mosa [3], which is involved in the biosynthesis of the rhizopine 3-o-methyl-scylo-inosamine. - Escherichia coli hypothetical protein yjhH. Two signature patterns for these enzymes were developed. The first one is centered on highly conserved region in the N-terminal part of these proteins. The second signature contains a lysine residue which has been shown, in Escherichia coli dapA [2], to be the one that forms a Schiff-base with the substrate.

Consensus pattern: [GSA]-[LIVM]-[LIVMFY]-x(2)-G-[ST]-[TG]-G-E-[GASNF]-x(6)-[EQ]

Consensus pattern: Y-[DNS]-[LIVMFA]-P-x(2)-[ST]-x(3)-[LIVMG]-x(13,14)-[LIVM]-x-[SGA]-[LIVMF]-K-[DEQAF]-[STAC] [K is involved in Schiff-base formation]-

[1] Kaneko T., Hashimoto T., Kumpaisal R., Yamada Y. J. Biol. Chem. 265:17451-17455(1990).

[2] Laber B., Gomis-Rueth F.-X., Romao M.J., Huber R. Biochem. J. 288:691-695(1992).

[3] Murphy P.J., Trenz S.P., Grzemski W., de Bruijn F.J., Schell J. J. Bacteriol. 175:5193-5204 (1993).

134. (DHodehase) Dihydroorotate dehydrogenase signatures

Dihydroorotate dehydrogenase (EC 1.3.3.1) (DHodehase) catalyzes the fourth step in the de novo biosynthesis of pyrimidine, the conversion of dihydroorotate into orotate. DHodehase is a ubiquitous FAD flavoprotein. In bacteria (gene pyrD), DHodease is located on the inner side of the cytosolic membrane. In some yeasts, such as in Saccharomyces cerevisiae (gene URA1), it is a cytosolic protein while in other eukaryotes it is found in the mitochondria [1]. The sequence of DHodease is rather well conserved and two signature patterns were developed specific to this enzyme. The first corresponds to a region in the N-terminal section of the enzyme while the second is located in the C-terminal section and seems to be part of the FAD-binding domain.

Consensus pattern[GS]-x(4)-[GK]-[GSTA]-[LIVFSTA]-[GT]-x(3)-[NOR]-x-G-[NHY]-x(2)-P-[RT]

Consensus pattern[LIVM](2)-[GSA]-x-G-G-[IV]-x-[STGDN]-x(3)-[ACV]-x(6)-G-A

[1] Nagy M., Lacroute F., Thomas D. Proc. Natl. Acad. Sci. U.S.A. 89:8966-8970(1992).

135. (DMRL_synthase) 6,7-dimethyl-8-ribityllumazine synthase

136. (DNA_methylase) C-5 cytosine-specific DNA methylases signatures

C-5 cytosine-specific DNA methylases (EC 2.1.1.73) (C5 Mtase) are enzymes that specifically methylate the C-5 carbon of cytosines in DNA [1,2,3]. Such enzymes are found in the proteins described below. - As a component of type II restriction-modification systems in prokaryotes and some bacteriophages. Such enzymes recognize a specific DNA sequence where they methylate a cytosine. In doing so, they protect DNA from cleavage by type II restriction enzymes that recognize the same sequence. The sequences of a large number of type II C-5 Mtases are known. - In vertebrates, there are a number of C-5 Mtases that methylate CpG dinucleotides. The sequence of the mammalian enzyme is known. C-5 Mtases share a number of short conserved regions. Two of them were selected. The first is centered around a conserved Pro-Cys dipeptide in which the cysteine has been shown [4] to be involved in the catalytic mechanism; it appears to form a covalent intermediate with the C6 position of cytosine. The second region is located at the C-terminal extremity in type-II enzymes

Consensus pattern: [DENKS]-x-[FLIV]-x(2)-[GSTC]-x-P-C-x(2)-[FYWLIM]-S [C is the active site residue]-

Consensus pattern: [RKOGTF]-x(2)-G-N-[STAG]-[LIVMF]-x(3)-[LIVMT]-x(3)-[LIVM]-x(3)-[LIVM]-

[1] Posfai J., Bhagwat A.S., Roberts R.J. Gene 74:261-263(1988).

[2] Kumar S., Cheng X., Klimasauskas S., Mi S., Posfai J., Roberts R.J., Wilson G.G. Nucleic Acids Res. 22:1-10(1994).

[3] Lauster R., Trautner T.A., Noyer-Weidner M. J. Mol. Biol. 206:305-312(1989).

[4] Chen L., McMillan A.M., Chang W., Ezak-Nipkay K., Lane W.S., Verdine G.L.
 Biochemistry 30:11018-11025(1991).

137. (DNaphotolyase) DNA photolyases class 2 signatures

Deoxyribodipyrimidine photolyase (EC 4.1.99.3) (DNA photolyase) [1,2] is a DNAREpair enzyme. It binds to UV-damaged DNA containing pyrimidine dimers and, upon absorbing a near-UV photon (300 to 500 nm), breaks the cyclobutane ring joining the two pyrimidines of the dimer. DNA photolyase is an enzyme that requires two chromophore-cofactors for its activity: a reduced FADH2 and either 5,10-methenyltetrahydrofolate (5,10-MTFH) or an oxidized 8-hydroxy-5-deazaflavin (8-HDF) derivative (F420). The folate or deazaflavin chromophore appears to function as an antenna, while the FADH2 chromophore is thought to be responsible for electron transfer. On the basis of sequence similarities[3] DNA photolyases can be grouped into two classes. The second class contains enzymes from Myxococcus xanthus, methanogenic archaeobacteria, insects, fish and marsupial mammals. It is not yet known what second cofactor is bound to class 2 enzymes. There are a number of conserved sequence regions in all known class 2 DNaphotolyases, especially in the C-terminal part. Two of these regions were selected as signature patterns.

Consensus pattern: F-x-E-E-x-[LIVM](2)-R-R-E-L-x(2)-N-F-

Consensus pattern: G-x-H-D-x(2)-W-x-E-R-x-[LIVM]-F-G-K-[LIVM]-R-[FY]-M-N-

[1] Sancar G.B., Sancar A. Trends Biochem. Sci. 12:259-261(1987).

[2] Jorns M.S. Biofactors 2:207-211(1990).

[3] Yasui A., Eker A.P.M., Yasuhira S., Yajima H., Kobayashi T., Takao M., Oikawa A.

EMBO J. 13:6143-6151(1994).

(DNaphotolyase2) DNA photolyases class 1 signatures

Deoxyribodipyrimidine photolyase (EC 4.1.99.3) (DNA photolyase) [1,2] is a DNA repair enzyme. It binds to UV-damaged DNA containing pyrimidine dimers and ,upon absorbing a near-UV photon (300 to 500 nm), breaks the cyclobutane ring joining the two pyrimidines of the dimer. DNA photolyase is an enzyme that requires two chromophore-cofactors for its activity: a reduced FADH2 and either 5,10-methenyltetrahydrofolate (5,10-MTFH) or an oxidized 8-hydroxy-5-deazaflavin (8-HDF) derivative (F420). The folate or deazaflavin

chromophore appears to function as an antenna, while the FADH2 chromophore is thought to be responsible for electron transfer. On the basis of sequence similarities[3] DNA photolyases can be grouped into two classes. The first class contains enzymes from Gram-negative and Gram-positive bacteria, the halophilic archaeobacteria *Halobacterium halobium*,
 5 fungi and plants. Class 1 enzymes bind either 5,10-MTHF (*E.coli*, fungi, etc.) or 8-HDF (*S.griseus*, *H.halobium*). This family also includes *Arabidopsis* cryptochromes 1 (CRY1) and 2 (CRY2), which are blue light photoreceptors that mediate blue light-induced gene expression. There are a number of conserved sequence regions in all known class 1 DNA photolyases, especially in the C-terminal part. Two of these regions were selected as
 10 signature patterns

Consensus pattern: T-G-x-P-[LIVM](2)-D-A-x-M-[RA]-x-[LIVM]-

Consensus pattern: [DN]-R-x-R-[LIVM](2)-x-[STA](2)-F-[LIVMFA]-x-K-x-L-x(2,3)-W-[KRO]-

[1] Sancar G.B., Sancar A. Trends Biochem. Sci. 12:259-261(1987).

[2] Jorns M.S. Biofactors 2:207-211(1990).

[3] Yasui A., Eker A.P.M., Yasuhira S., Yajima H., Kobayashi T., Takao M., Oikawa A. EMBO J. 13:6143-6151(1994).

[4] Lin C., Ahmad M., Cashmore A.R. Plant J. 10:893-902(1996).

138. (DNA_pol_A)

DNA polymerase family A signature

25 Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarities a number of DNA polymerases have been grouped together [1,2,3] under the designation of DNA polymerase family A. The polymerases that belong to this family are listed below.

- 30
- *Escherichia coli* and various other bacterial polymerase I (gene *polA*).
 - *Thermus aquaticus* Taq polymerase.
 - Bacteriophage sp01 polymerase.

- Bacteriophage sp02 polymerase.
 - Bacteriophage T5 polymerase.
 - Bacteriophage T7 polymerase.
 - Mycobacteriophage L5 polymerase.
- 5 - Yeast mitochondrial polymerase gamma (gene MIP1).

Five regions of similarity are found in all the above polymerases. One of these conserved regions, known as 'motif B' [1], is located in a domain which, in *Escherichia coli* polA, has been shown to bind deoxynucleotide triphosphate substrates; it contains a conserved tyrosine which has been shown, by photo-affinity labelling, to be in the active site; a conserved lysine, also part of this motif, can be chemically labelled, using pyridoxal phosphate. This conserved region was used as a signature for this family of DNA polymerases.

Consensus pattern R-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA] Sequences known to belong to this class detected by the pattern ALL.

- [1] Delarue M., Poch O., Todro N., Moras D., Argos P. *Protein Eng.* 3:461-467(1990).
- [2] Ito J., Braithwaite D.K. *Nucleic Acids Res.* 19:4045-4057(1991).
- [3] Braithwaite D.K., Ito J. *Nucleic Acids Res.* 21:787-802(1993).

139. DNA_pol_viral_C

DNA polymerase (viral) C-terminal domain

Number of members: 128

140. (DNA_topoisoII)

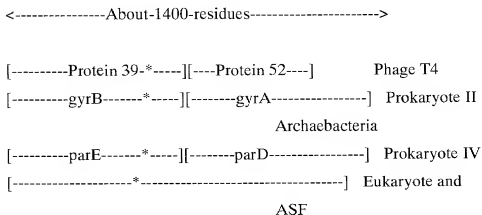
DNA topoisomerase II signature

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break.

Topoisomerase II is found in phages, archaeobacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits

(the product of genes 39, 52 and 60). In prokaryotes and in archaeobacteria the enzyme, known as DNA gyrase, consists of two subunits (genes *gyrA* and *gyrB* [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes *parC* and *parE*). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:



*: Position of the pattern.

As a signature pattern for this family of proteins, a region that contains a highly conserved pentapeptide was selected. The pattern is located in *gyrB*, in *parE*, and in protein 39 of phage T4 topoisomerase.

Consensus pattern[LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the pattern ALL.

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).

[2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).

[3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).

[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

141. (DSPc) Tyrosine specific protein phosphatases signature and profiles

Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation.

5 Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below: Soluble PTPases. - PTPN1 (PTP-1B). - PTPN2 (T-cell PTPase; TC-PTP). - PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1- like domain (see <PDOC00566>) and could act at junctions between the

10 membrane and cytoskeleton. - PTPN5 (STEP). - PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup. - PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HePTP). - PTPN8 (70Z-PEP). - PTPN9 (MEG2). - PTPN12 (PTP-G1; PTP-P19). - Yeast PTP1. - Yeast

15 PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway. - Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis. - Fission yeast pyp3 which contributes to the dephosphorylation of cdc2. - Yeast CDC14 which may be involved in chromosome segregation. - Yersinia virulence plasmid PTPases (gene yopH). - Autographa californica nuclear polyhedrosis virus 19 Kd PTPase. Dual specificity PTPases.

20 - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185. - DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues. - DUSP3 (VHR). - DUSP4 (HVH2). - DUSP5 (HVH3). - DUSP6 (Pyst1; MKP-3). - DUSP7 (Pyst2; MKP-X). - Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3. - Yeast YVH1.

25 - Vaccinia virus H1 PTPase; a dual specificity phosphatase. Receptor PTPases. Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The

30 cytoplasmic region generally contains two copies of the PTPase domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not. In the following table, the domain structure of known

receptor PTPases is shown: Extracellular Intracellular ----- Ig FN-3
 CAH MAM PTPaseLeukocyte common antigen (LCA) (CD45) 0 2 0 0 2Leukocyte antigen
 related (LAR) 3 8 0 0 2 Drosophila DLAR 3 9 0 0 2Drosophila DPTP 2 2 0 0 2PTP-alpha
 (LRP) 0 0 0 0 2PTP-beta 0 16 0 0 1PTP-gamma 0 1 1 0 2PTP-delta 0 >7 0 0 2 PTP-epsilon 0
 5 0 0 0 2PTP-kappa 1 4 0 1 2PTP-mu 1 4 0 1 2PTP-zeta 0 1 1 0 2PTPase domains consist of
 about 300 amino acids. There are two conserved cysteines, the second one has been shown to
 be absolutely required for activity. Furthermore, a number of conserved residues in its
 immediate vicinity have also been shown to be important. A signature pattern for PTPase
 domains was derived centered on the active site cysteine. There are three profiles for
 10 PTPases, the first one spans the complete domain and is not specific to any subtype. The
 second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily

Consensus pattern: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the
 active site residue]-

- [1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991).
- [2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992).
- [3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991).
- [4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989).
- 20 [5] Hunter T. Cell 58:1013-1016(1989).

142. (DUF10) Uncharacterized protein family UPF0076 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: -

- 25 Goat antigen UK114, a human homolog and the rat corresponding protein which is known as
 perchloric acid soluble protein (PSP1). PSP1 [2] may inhibit an initiation stage of cell-free
 protein synthesis. - Mouse heat-responsive protein HRSP12. - Yeast chromosome V
 hypothetical protein YER057c. - Yeast chromosome IX hypothetical protein YIL051c. -
 Caenorhabditis elegans hypothetical protein C23G10.2. - Escherichia coli hypothetical
 30 protein ycdK. - Escherichia coli hypothetical protein yhaR. - Escherichia coli hypothetical
 protein yjgF and HI0719, the corresponding Haemophilus influenzae protein. - Escherichia
 coli hypothetical protein yoaB. - Bacillus subtilis hypothetical protein yabJ. - Haemophilus
 influenzae hypothetical protein HI1627. - Helicobacter pylori hypothetical protein HP0944. -

Lactococcus lactis aldR. - Myxococcus xanthus dfrA. - Synechocystis strain PCC 6803
 hypothetical protein slr0709. - Rhizobium strain NGR234 symbiotic plasmid hypothetical
 protein y4sK. - Pyrococcus horikoshii hypothetical protein PH0854. These are small proteins
 of around 15 Kd whose sequence is highly conserved. As a signature pattern, a well conserved
 5 region located in the C-terminal part of these proteins was selected.

Consensus pattern: [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)-[GSAKR]-x-[LMVA]-
 x(5,8)-[LIVM]-E-[MI]-

[1] Bairoch A. Unpublished observations (1995).

[2] Oka T., Tsuji H., Noda C., Sakai K., Hong Y.-M., Suzuki I., Munoz S., Natori Y. J. Biol.
 Chem. 270:30060-30067(1995).

143. (DUF3) Domain of Unknown Function 3

Domain apparently occurring exclusively in eubacteria. Unknown
 function.

144. (DUF6) Integral membrane protein

This family includes many hypothetical membrane proteins of unknown function.
 Many of the proteins contain two copies of the aligned region.

145. (DUF7) Integral membrane protein

This family includes many hypothetical membrane proteins of unknown function.
Swiss:P14502 has been implicated in resistance to ethidium bromide.

146. (DapB) Dihydrodipicolinate reductase signature

Dihydrodipicolinate reductase (EC 1.3.1.26) catalyzes the second step in the biosynthesis of
 diaminopimelic acid and lysine, the NAD or NADP-dependent reduction of 2,3-
 dihydrodipicolinate into 2,3,4,5-tetrahydrodipicolinate. This enzyme is present in bacteria

(gene *dapB*) and higher plants. As a signature pattern the best conserved region in this enzyme was selected. It is located in the central section and is part of the substrate-binding region [1].

5 Consensus pattern: E-[IV]-x-E-x-H-x(3)-K-x-D-x-P-S-G-T-A-

[1] Scapin G., Blanchard J.S., Sacchettini J.C. *Biochemistry* 34:3502-3512(1995).

10 147. DedA family

This family combines the DedA related proteins and YIAN/YGIK family. Members of this family are not functionally characterised. These proteins contain multiple predicted transmembrane regions.

15

148. DegT/DnrJ/EryCI/StrS family

The members of this family exhibit some characteristics of the sensor protein of two-component signal transduction systems, however none of the members show any sequence similarity to these protein kinases. The members of this family do have the typical helix-turn-helix motif of DNA binding proteins.

20

[1] Stutzman-Engwall KJ, Otten SL, Hutchinson CR, *J Bacteriol* 1992;174:144-154.

149. (Desaturase) Fatty acid desaturases signatures

25 Fatty acid desaturases (EC 1.14.99.-) are enzymes that catalyze the insertion of a double bond at the delta position of fatty acids. There seems to be two distinct families of fatty acid desaturases which do not seem to be evolutionary related. Family I is composed of: -

Stearoyl-CoA desaturase (SCD) (EC 1.14.99.5) [1]. SCD is a key regulatory enzyme of unsaturated fatty acid biosynthesis. SCD introduces a cis double bond at the delta(9) position

30

of fatty acyl-CoA's such as palmitoleoyl- and oleoyl-CoA. SCD is a membrane-bound enzyme that is thought to function as a part of a multienzyme complex in the endoplasmic reticulum of vertebrates and fungi. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected, this region is rich in histidine residues

and in aromatic residues. Family 2 is composed of: - Plants stearoyl-acyl-carrier-protein desaturase (EC 1.14.99.6) [2], these enzymes catalyze the introduction of a double bond at the delta(9) position of stearoyl-ACP to produce oleoyl-ACP. This enzyme is responsible for the conversion of saturated fatty acids to unsaturated fatty acids in the synthesis of vegetable oils. - Cyanobacteria desA [3] an enzyme that can introduce a second cis double bond at the delta(12) position of fatty acid bound to membranes glycerolipids. DesA is involved in chilling tolerance; the phase transition temperature of lipids of cellular membranes being dependent on the degree of unsaturation of fatty acids of the membrane lipids. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected.

Consensus pattern: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y-

Consensus pattern: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]-[DE]-

[1] Kaestner K.H., Ntambi J.M., Kelly T.J. Jr., Lane M.D. J. Biol. Chem. 264:14755-14761(1989).

[2] Shanklin J., Somerville C.R. Proc. Natl. Acad. Sci. U.S.A. 88:2510-2514(1991).

[3] Wada H., Gombos Z., Murata N. Nature 347:200-203(1990).

150. Dihydroorotase signatures

Dihydroorotase (EC 3.5.2.3) (DHOase) catalyzes the third step in the de novo biosynthesis of pyrimidine, the conversion of ureidosuccinic acid (N-carbamoyl-L-aspartate) into dihydroorotate. Dihydroorotase binds a zinc ion which is required for its catalytic activity [1].

In bacteria, DHOase is a dimer of identical chains of about 400 amino-acid residues (gene pyrC). In higher eukaryotes, DHOase is part of a large multi-functional protein known as 'rudimentary' in Drosophila and CAD in mammals and which catalyzes the first three steps of pyrimidine biosynthesis [2]. The DHOase domain is located in the central part of this polypeptide. In yeasts, DHOase is encoded by a monofunctional protein (gene URA4).

However, a defective DHOase domain [3] is found in a multifunctional protein (gene URA2) that catalyzes the first two steps of pyrimidine biosynthesis. The comparison of DHOase sequences from various sources shows [4] that there are two highly conserved regions. The first located in the N-terminal extremity contains two histidine residues

suggested [3] to be involved in binding the zinc ion. The second is found in the C-terminal part. Signature patterns for both regions have been developed. Allantoinase (EC 3.5.2.5) is the enzyme that hydrolyzes allantoin into allantoate. In yeast (gene DAL1) [5], it is the first enzyme in the allantoin degradation pathway; in amphibians [6] and fish it catalyzes the second step in the degradation of uric acid. The sequence of allantoinase is evolutionary related to that of DHOases.

Consensus pattern: D-[LIVMFYWSAP]-H-[LIVA]-H-[LIVF]-[RN]-x-[PGANF] [The two H's are probable zinc ligands]-

Consensus pattern: [GA]-[ST]-D-x-A-P-H-x(4)-K-

[1] Brown D.C., Collins K.D. J. Biol. Chem. 266:1597-1604(1991).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[3] Souciet J.-L., Nagy M., Le Gouar M., Lacroute F., Potier S. Gene 79:59-70(1989).

[4] Guyonvarch A., Nguyen-Juilleret M., Hubert J.-C., Lacroute F. Mol. Gen. Genet. 212:134-141(1988).

[5] Buckholz R.G., Cooper T.G. Yeast 7:913-923(1991).

[6] Hayashi S., Jain S., Chu R., Alvares K., Xu B., Erfurth F., Usuda N., Rao M.S., Reddy S.K., Noguchi T., Reddy J.K., Yeldandi A.Y. J. Biol. Chem. 269:12269-12276(1994).

151. dnaJ domains signatures and profile

The prokaryotic heat shock protein dnaJ interacts with the chaperone hsp70-like dnaK protein [1]. Structurally, the dnaJ protein consists of an N-terminal conserved domain (called 'J' domain) of about 70 amino acids, a glycine-rich region ('G' domain') of about 30 residues, a central domain containing four repeats of a CXXCXGXG motif ('CRR' domain) and a C-terminal region of 120 to 170 residues. Such a structure is shown in the following schematic representation:

```

+-----+-----+-----+-----+-----+-----+-----+-----+ | N-terminal | |
Gly-R | | CXXCXGXG | C-terminal | +-----+-----+-----+-----+-----+-----+
-----+
```

It has been shown [2] that the 'J' domain as well as the 'CRR' domain are also found in other prokaryotic and eukaryotic proteins which are listed below.

a) Proteins containing both a 'J' and a 'CRR' domain:

- Yeast protein MAS5/YDJ1 which seems to be involved in mitochondrial protein import.
- Yeast protein MDJ1, involved in mitochondrial biogenesis and protein folding.
- Yeast protein SCJ1, involved in protein sorting.
- Yeast protein XDJ1.
- Plants dnaJ homologs (from leek and cucumber).
- Human HDJ2, a dnaJ homolog of unknown function.
- Yeast hypothetical protein YNL077w.

b) Proteins containing a 'J' domain without a 'CRR' domain:

- Rhizobium fredii nolC, a protein involved in cultivar-specific nodulation of soybean.
- Escherichia coli cbpA [3], a protein that binds curved DNA.
- Yeast protein SEC63/NPL1, important for protein assembly into the endoplasmic reticulum and the nucleus.
- Yeast protein SIS1, required for nuclear migration during mitosis.
- Yeast protein CAJ1.
- Yeast hypothetical protein YFR041c.
- Yeast hypothetical protein YIR004w.
- Yeast hypothetical protein YJL162c.
- Plasmodium falciparum ring-infected erythrocyte surface antigen (RESA). RESA, whose function is not known, is associated with the membrane skeleton of newly invaded erythrocytes.
- Human HDJ1.
- Human HSJ1, a neuronal protein.
- Drosophila cysteine-string protein (csp).

A signature pattern for the 'J' domain was developed, based on conserved positions in the C-terminal half of this domain. A pattern for the 'CRR' domain, based on the first two copies of that motif was also developed. A profile for the 'J' domain was also developed.

Consensus pattern: [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI]-

Consensus pattern: C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G-

[1] Cyr D.M., Langer T., Douglas M.G. Trends Biochem. Sci. 19:176-181(1994).

[2] Bork P., Sander C., Valencia A., Bukau B. Trends Biochem. Sci. 17:129-129(1992).

[3] Ueguchi C., Kaneda M., Yamada H., Mizuno T. Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).

152.

153. Dwarfing

This family known as the dwarfins also includes the drosophila protein MAD. The N-terminus of MAD can bind to DNA [2].

[1] Yingling JM, Das P, Savage C, Zhang M, Padgett RW, Wang XF, Proc Natl Acad Sci U S A 1996;93:8940-8944. [2] Kim J, Johnson K, Chen HJ, Carroll S, Laughon A, Nature 1997;388:304-308.

154. Dynein light chain type 1 signature

Dynein is a multisubunit microtubule-dependent motor enzyme that acts as the force generating protein of eukaryotic cilia and flagella. The cytoplasmic isoform of dynein acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules. Dynein is composed of a number of ATP-binding large subunits, intermediate size subunits and small subunits. Among the small subunits, there is a family [1,2] of highly conserved proteins which consist of: - Chlamydomonas reinhardtii flagellar outer arm dynein 8 Kd and 11 Kd light chains. - Higher eukaryotes cytoplasmic dynein light chain 1. - Yeast cytoplasmic dynein light chain 1 (gene DYN2 or SLC1). - Caenorhabditis elegans hypothetical dynein light chains M18.2 and T26A5.9. These proteins have from 89 to 120 amino acids. As a signature pattern, A highly conserved region was selected.

Consensus pattern: H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E-

- [1] King S.M., Patel-King R.S. J. Biol. Chem. 270:11445-11452(1995).
- [2] Dick T., Ray K., Salz H.K., Chia W. Mol. Cell. Biol. 16:1966-1977(1996).

5

155. dUTPase

dUTPase hydrolyzes dUTP to dUMP and pyrophosphate.

- [1] Cedergren-Zeppezauer ES, Larsson G, Nyman PO, Dauter Z, Wilson KS, Nature 10 1992;355:740-743. [2] Mol CD, Harris JM, McIntosh EM, Tainer JA, Structure 1996;4:1077-1092.

156. (dCMP cyt deam) Cytidine and deoxycytidylate deaminases zinc-binding region signature

Cytidine deaminase (EC [3.5.4.5](#)) (cytidine aminohydrolase) catalyzes the hydrolysis of cytidine into uridine and ammonia while deoxycytidylatedeaminase (EC [3.5.4.12](#)) (dCMP deaminase) hydrolyzes dCMP into dUMP. Both enzymes are known to bind zinc and to require it for their catalytic activity[1,2]. These two enzymes do not share any sequence similarity with the exception of a region that contains three conserved histidine and cysteine residues which are thought to be involved in the binding of the catalytic zincion. Such a region is also found in other proteins [3,4]: - Yeast cytosine deaminase (EC [3.5.4.1](#)) (gene FCY1) which transforms cytosine into uracil. - Mammalian apolipoprotein B mRNA editing protein, responsible for the postranscriptional editing of a CAA codon into a UAA (stop) codon in the APOB mRNA. - Riboflavin biosynthesis protein ribG, which converts 2,5-diamino-6- (ribosylamino)-4(3H)-pyrimidinone 5'-phosphate into 5-amino-6- (ribosylamino)-2,4(1H,3H)-pyrimidinedione 5'-phosphate. - *Bacillus cereus* blasticidin-S deaminase (EC [3.5.4.23](#)), which catalyzes the deamination of the cytosine moiety of the antibiotics blasticidin S, cytomycin and acetylblasticidin S. - *Bacillus subtilis* protein comEB. This protein is required for the binding and uptake of transforming DNA. - *Bacillus subtilis* hypothetical protein yaaJ. - *Escherichia coli* hypothetical protein yfhC. - Yeast hypothetical protein YJL035c. A signature pattern for this zinc-binding region was derived.

Consensus pattern: [CH]-[AGV]-E-x(2)-[LIVMFGAT]-[LIVM]-x(17,33)-P-C-x(2,8)-C-x(3)-[LIVM] [The C's and H are zinc ligands]

- [1] Yang C., Carlow D., Wolfenden R., Short S.A. Biochemistry 31:4168-4174(1992).
 5 [2] Moore J.T., Silversmith R.E., Maley G.F., Maley F. J. Biol. Chem. 268:2288-2291(1993).
 [3] Reizer J., Buskirk S., Bairoch A., Reizer A., Saier M.H. Jr. Protein Sci. 3:853-856(1994).
 [4] Bhattacharya S., Navaratnam N., Morrison J.R., Scott J., Taylow W.R. Trends Biochem. Sci. 19:105-106(1994).

10

157. Dehydrins signatures

A number of proteins are produced by plants that experience water-stress. Water-stress takes place when the water available to a plant falls below a critical level. The plant hormone abscisic acid (ABA) appears to modulate the response of plant to water-stress. Proteins that are expressed during water-stress are called dehydrins [1,2] or LEA group 2 proteins [3]. The proteins that belong to this family are listed below.

- Arabidopsis thaliana XERO 1, XERO 2 (LTI30), RAB18, ERD10 (LTI45) ERD14 and COR47.
- Barley dehydrins B8, B9, B17, and B18.
- Cotton LEA protein D-11.
- Craterostigma plantagineum dessication-related proteins A and B.
- Maize dehydrin M3 (RAB-17).
- Pea dehydrins DHN1, DHN2, and DHN3.
- Radish LEA protein.
- Rice proteins RAB 16B, 16C, 16D, RAB21, and RAB25.
- Tomato TAS14.
- Wheat dehydrin RAB 15 and cold-shock protein cor410, cs66 and cs120.

Dehydrins share a number of structural features. One of the most notable features is the presence, in their central region, of a continuous run of five to nine serines followed by a cluster of charged residues. Such a region has been found in all known dehydrins so far with the exception of pea dehydrins. A second conserved feature is the presence of two copies of lysine-rich octapeptide; the first copy is located just after the cluster of charged residues that follows the poly-serine region and the second copy is found at the C-terminal extremity. Signature patterns for both regions were derived.

30

Consensus pattern: S(5)-[DE]-x-[DE]-G-x(1,2)-G-x(0,1)-[KR](4

Consensus pattern: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G-

- 5 [1] Close T.J., Kortt A.A., Chandler P.M. *Plant Mol. Biol.* 13:95-108(1989).
 [2] Robertson M., Chandler P.M. *Plant Mol. Biol.* 19:1031-1044(1992).
 [3] Dure L. III, Crouch M., Harada J., Ho T.-H. D., Mundy J., Quatrano R., Thomas T., Sung Z.R. *Plant Mol. Biol.* 12:475-486(1989).

10

158. (deoR) Bacterial regulatory proteins, deoR family signature

The many bacterial transcription regulation proteins which bind DNA through a helix-turn-helix' motif can be classified into subfamilies on the basis of sequence similarities. One of these subfamilies groups the following proteins[1,2]: - accR, the *Agrobacterium tumefaciens* plasmid pTiC58 repressor of opine catabolism and conjugal transfer. - agaR, the *Escherichia coli* aga operon putative repressor. - deoR, the *Escherichia coli* deoxyribose operon repressor. - fucR, the *Escherichia coli* L-fucose operon activator. - gatR, the *Escherichia coli* galactitol operon repressor. - glpR, the *Escherichia coli* glycerol-3-phosphate regulon repressor. - gutR (or srlR), the *Escherichia coli* glucitol operon repressor. - ioIR, from *Bacillus subtilis*. - lacR, the streptococci lactose phosphotransferase system repressor. - spoIIID, the *Bacillus subtilis* transcription regulator of the sigK gene. - yfjR, an *Escherichia coli* hypothetical protein. - ygbI, an *Escherichia coli* hypothetical protein. - yihW, an *Escherichia coli* hypothetical protein. - yjfQ, an *Escherichia coli* hypothetical protein. - yjhJ, an *Escherichia coli* hypothetical protein. The 'helix-turn-helix' DNA-binding motif of these proteins is located in the N-terminal part of the sequence. The pattern used to detect these proteins starts fourteen residues before the HTH motif and ends one residue after it.

Consensus pattern: R-x(3)-[LIVM]-x(3)-[LIVM]-x(16,17)-[STA]-x(2)-T-[LIVMA]-[RH]-[KRNA]-D-[LIVMF]-

30

[1] von Bodman S., Hayman G.T., Farrand S.K. *Proc. Natl. Acad. Sci. U.S.A.* 89:643-647(1992).

[2] Bairoch A. Unpublished observations (1993).

159. dsrm

Double-stranded RNA binding motif

5

[1] Burd CG, Dreyfuss G; Medline: 94310455, Conserved structures and diversity of functions of RNA-binding proteins. Science 1994;265:615-621.

Sequences gathered for seed by HMM_iterative_training Putative motif shared by proteins that bind to dsRNA. At least some DSRM proteins seem to bind to specific RNA targets. Exemplified by Staufen, which is involved in localization of at least five different mRNAs in the early Drosophila embryo. Also by interferon-induced protein kinase in humans, which is part of the cellular response to dsRNA.

10

15

Number of members: 116

160. Dynamin family signature

Dynamin [1,2] is a microtubule-associated force-producing protein of 100 Kd which is involved in the production of microtubule bundles and which is able to bind and hydrolyze GTP. Dynamin is structurally related to the following proteins: - Drosophila shibire protein (gene shi) [3]. Shibire is, very probably, the Drosophila cognate of mammalian dynamin. It seems to provide the motor for vesicular transport during endocytosis. - Yeast vacuolar sorting protein VPS1 (or SPO15) [4], a protein which could also be involved in microtubule-associated motility. - Yeast protein MGM1 [5], which is required for mitochondrial genome maintenance. - Yeast protein DNM1, which is involved in endocytosis. - Interferon induced Mx proteins [6,7]. Interferon alpha or beta induce the synthesis of a family of closely related proteins. Most of these proteins are known to confer resistance to influenza viruses and/or rhabdoviruses on transfected mammalian cell in culture. The three motifs found in all GTP-binding proteins are located in the N-terminal part of these proteins. The signature pattern that was developed for these proteins is based on a highly conserved region downstream of the ATP/GTP-binding motif 'A' (P-loop) (see <PDOC00017>).-

20

25

30

Consensus pattern: L-P-[RK]-G-[STN]-[GN]-[LIVM]-V-T-R-

- [1] Vallee R.B., Shpetner H.S. Annu. Rev. Biochem. 59:909-932(1990).
 [2] Obar R.A., Collins C.A., Hammarback J.A., Shpetner H.S., Vallee R.B. Nature 347:256-
 261(1990).
 [3] van der Blik A., Meyerowitz E.M. Nature 351:411-414(1991).
 [4] Rothman J.H., Raymond C.K., Gilbert T., O'Hara P.J., Stevens T.H. Cell 61:1063-1074(1990).
 [5] Jones B.A., Fangman W.L. Genes Dev. 6:380-389(1992).
 [6] Arnheiter H., Meier E. New Biol. 2:851-857(1990).
 [7] Staeheli P., Pitossi F., Pavlovic J. Trends Cell Biol. 3:268-272(1993).

161. (dynamin_2) Dynamamin central region

This region lies between the GTPase domain, see dynamamin, and the pleckstrin homology (PH) domain.

162. E1-E2 ATPases phosphorylation site

E1-E2 ATPases (also known as P-type) are cation transport ATPases which form an aspartyl phosphate intermediate in the course of ATP hydrolysis. ATPases which belong to this family are listed below [1,2,3]. - Fungal and plant plasma membrane (H+) ATPases [reviewed in 4]. - Vertebrate (Na+, K+) ATPases (sodium pump) [reviewed in 5,6]. - Gastric (K+, H+) ATPases (proton pump). - Calcium (Ca++) ATPases (calcium pump) from the sarcoplasmic reticulum (SR), the endoplasmic reticulum (ER) and the plasma membrane. - Copper (Cu++) ATPases (copper pump) which are involved in two human genetic disorders: Menkes syndrome and Wilson disease [7]. - Bacterial potassium (K+) ATPases. - Bacterial cadmium efflux (Cd++) ATPases [reviewed in 8]. - Bacterial magnesium (Mg++) ATPases. - A probable cation ATPase from Leishmania. - fixI, a probable cation ATPase from Rhizobium meliloti, involved in nitrogen fixation. The region around the phosphorylated aspartate residue is perfectly conserved in all these ATPases and can be used as a signature pattern.

Consensus pattern: D-K-T-G-T-[LI]-[TI] [D is phosphorylated]

- [1] Green N.M., McLennan D.H. Biochem. Soc. Trans. 17:819-822(1989).
- [2] Green N.M. Biochem. Soc. Trans. 17:970-972(1989).
- [3] Fagan M.J., Saier M.H. Jr. J. Mol. Evol. 38:57-99(1994).
- 5 [4] Serrano R. Biochim. Biophys. Acta 947:1-28(1988).
- [5] Fambrough D.M. Trends Neurosci. 11:325-328(1988).
- [6] Sweadner K.J. Biochim. Biophys. Acta 988:185-220(1989).
- [7] Bull P.C., Cox D.W. Trends Genet. 10:246-251(1994).
- [8] Silver S., Nucifora G., Chu L., Misra T.K. Trends Biochem. Sci. 14:76-80(1989).

163. E1_N

E1 Protein, N terminal domain

Number of members: 90

164. (E1_dehydrog) Dehydrogenase E1 component

This family uses thiamine pyrophosphate as a cofactor. This family includes pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and 2-oxoisovalerate dehydrogenase.

165. (ECH) Enoyl-CoA hydratase/isomerase signature

Enoyl-CoA hydratase (EC 4.2.1.17) (ECH) [1] and 3-2trans-enoyl-CoA isomerase (EC 5.3.3.8) (ECI) [2] are two enzymes involved in fatty acid metabolism. ECH catalyzes the hydration of 2-trans-enoyl-CoA into 3-hydroxyacyl-CoA and ECI shifts the 3- double bond of the intermediates of unsaturated fatty acid oxidation to the 2-trans position. Most eukaryotic cells have two fatty-acid beta-oxidation systems, one located in mitochondria and the other in peroxisomes. In mitochondria, ECH and ECI are separate yet structurally related monofunctional enzymes. Peroxisomes contain a trifunctional enzyme [3] consisting of an N-terminal domain that bears both ECH and ECI activity, and a C-terminal domain responsible for 3-hydroxyacyl-CoA dehydrogenase (HCDH) activity. In *Escherichia coli* (gene *fadB*) and *Pseudomonas fragi* (gene *faoA*), ECH and ECI are also part of a multifunctional enzyme which contains both a HCDH and a 3-hydroxybutyryl-CoA epimerase domain [4]. A number

of other proteins have been found to be evolutionary related to the ECH/ECI enzymes or domains: - 3-hydroxybutyryl-coa dehydratase (EC 4.2.1.55) (crotonase), a bacterial enzyme involved in the butyrate/butanol-producing pathway. - Naphthoate synthase (EC 4.1.3.36) (DHNA synthetase) (gene *menB*) [5], a bacterial enzyme involved in the biosynthesis of menaquinone (vitamin K2). DHNA synthetase converts O-succinyl-benzoyl-CoA (OSB-CoA) to 1,4-dihydroxy- 2-naphthoic acid (DHNA). - 4-chlorobenzoate dehalogenase (EC 3.8.1.6) [6], a *Pseudomonas* enzyme which catalyzes the conversion of 4-chlorobenzoate-CoA to 4-hydroxybenzoate-CoA. - A *Rhodobacter capsulatus* protein of unknown function (ORF257) [7]. - *Bacillus subtilis* putative polyketide biosynthesis proteins *pksH* and *pksI*. - *Escherichia coli* carnitine racemase (gene *caiD*) [8]. - *Escherichia coli* hypothetical protein *ygfG*. - Yeast hypothetical protein YDR036c. As a signature pattern for these enzymes, a conserved region rich in glycine and hydrophobic residues was selected.

Consensus pattern: [LIVM]-[STA]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-[DQHP]-[LIVMFY]-

- [1] Minami-Ishii N., Taketani S., Osumi T., Hashimoto T. *Eur. J. Biochem.* 185:73-78(1989).
- [2] Mueller-Newen G., Stoffel W. *Biol. Chem. Hoppe-Seyler* 372:613-624(1991).
- [3] Palosaari P.M., Hiltunen J.K. *J. Biol. Chem.* 265:2446-2449(1990).
- [4] Nakahigashi K., Inokuchi H. *Nucleic Acids Res.* 18:4937-4937(1990).
- [5] Driscoll J.R., Taber H.W. *J. Bacteriol.* 174:5063-5071(1992).
- [6] Babbitt P.C., Kenyon G.L., Matin B.M., Charest H., Sylvestre M., Scholten J.D., Chang K.-H., Liang P.-H., Dunaway-Mariano D. *Biochemistry* 31:5594-5604(1992).
- [7] Beckman D.L., Kranz R.G. *Gene* 107:171-172(1991).
- [8] Eichler K., Bourgis F., Buchet A., Kleber H.-P., Mandrand-Berthelot M.-A. *Mol. Microbiol.* 13:775-786(1994).

166. (EF1BD) Elongation factor 1 beta/beta'/delta chain signatures

Eukaryotic elongation factor 1 (EF-1) is responsible for the GTP-dependent binding of aminoacyl-tRNAs to the ribosomes [1]. EF-1 is composed of four subunits: the alpha chain which binds GTP and aminoacyl-tRNAs, the gamma chain that probably plays a role in

anchoring the complex to other cellular components and the beta and delta (or beta') chains. The beta and delta chains are highly similar proteins that both stimulate the exchange of GDP bound to the alpha chain for GTP [2]. The beta and delta chains are hydrophilic proteins of around 23 to 31 Kd. Their C-terminal part seems important for the nucleotide exchange activity, while the N-terminal section is probably involved in the interaction with the gamma chain. Two signature patterns for this family of proteins were developed. The first corresponds to an acidic region in the central section; the second, to the C-terminal extremity of these proteins

Consensus pattern: [DE]-[DEG]-[DE](2)-[LIVMF]-D-L-F-G-

Consensus pattern: [IV]-Q-S-x-D-[LIVM]-x-A-[FWM]-[NQ]-K-[LIVM]-

[1] Riis B., Rattan I.S., Clark B.F.C., Merrick W.C. Trends Biochem. Sci. 15:420-424(1990).

[2] van Damme H.T.F., Amons R., Karssies R., Timmers C.J., Janssen G.M.C., Moeller W. Biochim. Biophys. Acta 1050:241-247(1990).

167. (EF1G_domain) Elongation factor 1 gamma, conserved domain

168. (EFG_C) Elongation factor G C-terminus

This family is always found associated with GTP_EFTU. This family includes the carboxyl terminal regions of Elongation factor G, elongation factor 2 and some tetracycline resistance proteins.

169. (EFP) Elongation factor P signature

Elongation factor P (EF-P) [1] is a prokaryotic protein translation factor required for efficient peptide bond synthesis on 70S ribosomes from fMet-tRNA^{fMet}. EF-P is a protein of 21 Kd.

It is evolutionary related to yeiP, an hypothetical protein from Escherichia coli. As a signature pattern, a conserved region located in the C-terminal part of these proteins was selected.

203

Consensus pattern: K-x-[AV]-x(4)-G-x(2)-[LIV]-x-V-P-x(2)-[LIV]-x(2)-G-

[1] Aoki H., Adams S.-L., Turner M.A., Ganoza M.C. Biochimie 79:7-11(1997).

5

170. (EF TS) Elongation factor Ts signatures

In prokaryotes elongation factor Ts (EF-Ts) is a component of the elongation cycle of protein biosynthesis. It associates with the EF-Tu.GDP complex and induces the exchange of GDP to GTP, it remains bound to the aminoacyl-tRNA.EF-Tu.GTP complex up to the GTP

10 hydrolysis stage on the ribosome [1].EF-Ts is also a component of the chloroplast protein biosynthetic machinery and is encoded in the genome of some algal chloroplast [2]. It is also present in mitochondria [3]. As signature patterns for EF-Ts, two conserved regions located in the N-terminal part of the protein have been selected.

15 Consensus pattern: L-R-x(2)-T-[GSDNQ]-x-[GS]-[LIVMF]-x(0,1)-[DENKAC]-x-K-[KRNEQS]-A-L-

Consensus pattern: E-[LIVM]-[NV]-[SCV]-[QE]-T-D-F-V-[SA]-[KRN]-

[1] Bubunenko M.G., Kireeva M.L., Gudkov A.T. Biochimie 74:419-425(1992).

20 [2] Kostrzewa M., Zetsche K. Plant Mol. Biol. 23:67-76(1993).

[3] Xin H., Worriax V.L., Burkhart W.A., Spremulli L.L. J. Biol. Chem. 270:17243-17249(1995).

25 171. (EMP24_GP25L) emp24/gp25L/p24 family

Members of this family are implicated in bringing cargo forward from the ER and binding to coat proteins by their cytoplasmic domains. Number of members: 30

Paccaud JP, Thomas DY, Bergeron JJ, Nilsson T, J Cell Biol 1998;140:751-765.

30

172. ENV_polyprotein

ENV polyprotein (coat polyprotein)

Number of members: 224

173. (ERG4_ERG24) Ergosterol biosynthesis ERG4/ERG24 family signatures

Two fungal enzymes involved in ergosterol biosynthesis and which act by reducing double bonds in precursors of ergosterol have been shown to be evolutionary related [1]. These are C-14 sterol reductase (gene ERG24 in budding yeast and *erg3* in *Neurospora Crassa*) and C-24(28) sterol reductase (gene ERG4 in budding yeast and *sts1* in fission yeast). Their sequences are also highly related to that of chicken lamin B receptor, which is thought to anchor the lamina to the inner nuclear membrane. These proteins are highly hydrophobic and seem to contain seven or eight transmembrane regions. As signature patterns, two conserved regions were selected. The first one is apparently located in a loop between the fourth and fifth transmembrane regions and the second is in the C-terminal section.

Consensus pattern: G-x(2)-[LIVM]-[YH]-D-x-[FYW]-x-G-x(2)-L-N-P-R -

Consensus pattern: [LIVM](2)-H-R-x(2)-R-D-x(3)-C-x(2)-K-Y-G-

[1] Lai M.H., Bard M., Pierson C.A., Alexander J.F., Goebel M., Carter G.T., Kirsch D.R. Gene 140:41-49(1994).

174. (ERM) Ezrin/radixin/moesin family

This family of proteins contain a band 4.1 domain (Band_41), at their amino terminus. This family represents the rest of these proteins.

[1] Yonemura S, Hirao M, Doi Y, Takahashi N, Kondo T, Tsukita S, J Cell Biol 1998;140:885-895.

175. ER lumen protein retaining receptor signatures

Proteins that reside in the lumen of the endoplasmic reticulum (ER) contain aC-terminal tetrapeptide (generally K-D-E-L or H-D-E-L) that serves as a signal for their retrieval (retrograde transport) from subsequent compartments of the secretory pathway. The signal is recognized by a receptor molecule that is believed to cycle between the cis side of the Golgi apparatus and the ER [1]. This protein is known as the ER lumen protein retaining receptor or

also as the 'KDEL receptor'. It has been characterized in a variety of species, including fungi (gene ERD2), plants, Plasmodium, Drosophila and mammals. In mammals two highly related forms of the receptor are known. Structurally, the receptor is a protein of about 220 residues that seems to contain seven transmembrane regions [2]. The N-terminal part (3 residues) is oriented toward the lumen while the C-terminal tail (about 12 residues) is cytoplasmic. There are three luminal and three cytoplasmic loops. Two signature patterns for these receptors were developed. The first pattern corresponds to the C-terminal half of the first cytoplasmic loop as well as most of the second transmembrane domain. The second pattern is a perfectly conserved decapeptide that corresponds to the central part of the fifth transmembrane domain.

Consensus pattern: G-I-S-x-[KR]-x-Q-x-L-[FY]-x-[LIV](2)-F-x(2)-R-Y-
Consensus pattern: L-E-[SA]-V-A-I-[LM]-P-Q-L-

- [1] Pelham H.R.B. Curr. Opin. Cell Biol. 3:585-591(1991).
- [2] Townsley F.M., Wilson D.W., Pelham H.R.B. EMBO J. 12:2821-2829(1993).

176. (ETF_beta) Electron transfer flavoprotein beta-subunit signature

The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria. The beta subunit of ETF is a protein of about 28 Kd which is structurally related to the bacterial nitrogen fixation protein fixA which could play a role in a redox process and feed electrons to ferredoxin. Other related proteins are: - Escherichia coli hypothetical protein ydiQ. - Escherichia coli hypothetical protein ygcR. As a signature pattern for these proteins, a conserved region which is located in the central section was selected.

Consensus pattern: [IVA]-x-[KR]-x(2)-[DE]-[GD]-[GDE]-x(1,2)-[EQ]-x-[LIV]- x(4)-P-x-[LIVM](2)-[TAC]-

- [1] Finocchiario G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990).

[2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).

177. Endonuclease III signatures

5 Escherichia coli endonuclease III (EC 4.2.99.18) (gene nth) [1] is a DNA repair enzyme that acts both as a DNA N-glycosylase, removing oxidized pyrimidines from DNA, and as an apurinic/apyrimidinic (AP) endonuclease, introducing a single-strand nick at the site from which the damaged base was removed. Endonuclease III is an iron-sulfur protein that binds a single 4Fe-4S cluster. The 4Fe-4S cluster does not seem to be important for catalytic activity, but is probably involved in the proper positioning of the enzyme along the DNA strand [2]. Endonuclease III is evolutionary related to the following proteins: - Fission yeast endonuclease III homolog (gene nth1) [3]. - Escherichia coli and related protein DNA repair protein mutY, which is an adenine glycosylase. MutY is a larger protein (350 amino acids) than endonuclease III (211 amino acids). - Micrococcus luteus ultraviolet N-glycosylase/AP lyase which initiates repair at cis-syn pyrimidine dimers. - ORF10 in plasmid pFV1 of the thermophilic archaeobacteria Methanobacterium thermoformicum [4]. Restriction methylase m.MthTI, which is encoded by this plasmid, generates 5-methylcytosine which is subject to deamination resulting in G-T mismatches. This protein could correct these mismatches. - Yeast hypothetical protein YAL015c. - Fission yeast hypothetical protein SpAC26A3.02. - Caenorhabditis elegans hypothetical protein R10E4.5. - Methanococcus jannaschii hypothetical protein MJ0613. The 4Fe-4S cluster is bound by four cysteines which are all located in a 17 amino acid region at the C-terminal end of endonuclease III. A similar region is also present in the central section of mutY and in the C-terminus of ORF10 and of the Micrococcus UV endonuclease. The 4Fe-4S cluster region does not exist in YAL015c. Two signature patterns for these proteins were developed: the first corresponds to the core of the iron-sulfur binding domain, the second corresponds to the best conserved region in the catalytic core of these enzymes.

Consensus pattern: C-x(3)-[KRS]-P-[KRAGL]-C-x(2)-C-x(5)-C [The four C's are 4Fe-4S ligands]-

Consensus pattern: [GST]-x-[LIVMF]-P-x(5)-[LIVMW]-x(2,3)-[LI]-[PAS]-G-V-[GA]-x(3)-[GAC]-x(3)-[LIVM]-x(2)-[SALV]-[LIVMFYW]-[GANK]-

[1] Kuo C.-F., McRae D., Fisher C.L., O'Handley S.F., Cunningham R.P., Tainer J.A. Science 258:434-440(1992).

[2] Thomson A.J. Curr. Biol. 3:173-174(1993).

[3] Roldan-Arjona T., Anselmino C., Lindahl T. Nucleic Acids. Res. 3307-3312(1996).

5 [4] Noelling J., van Eeden F.J.M., Eggen R.I.L., de Vos W.M. Nucleic Acids Res. 20:6501-6507(1992).

178. (Epimerase) NAD dependent epimerase/dehydratase family

10 This family of proteins utilize NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions.

[1] Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM, Biochemistry 1997;36:6294-6304.

179. Exonuclease

15 This family includes a variety of exonuclease proteins, such as ribonuclease T and the epsilon subunit of DNA polymerase III.

[1] Koonin EV, Deutscher MP, Nucleic Acids Res 1993;21:2521-2522.

180. ENTH

ENTH domain

25 [1] Kay BK, Yamabhai M, Wendland B, Emr SD; Medline: 99156083, Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery. Protein Sci 1999;8:435-438.

30 The ENTH (Epsin N-terminal homology) domain is found in proteins involved in endocytosis and cytoskeletal machinery. The function of the ENTH domain is unknown.

Number of members: 29

181. (eIF-1A) Eukaryotic initiation factor 1A signature

Eukaryotic translation initiation factor 1A (eIF-1A) [1] (formerly known as eIF-4C) is a protein that seems to be required for maximal rate of protein biosynthesis. It enhances ribosome dissociation into subunits and stabilizes the binding of the initiator Met-tRNA to 40S ribosomal subunits. eIF-1A is a hydrophilic protein of about 15 to 17 Kd. Archaeobacteria also seem to possess a eIF-1A homolog. As a signature pattern, a conserved region in the central section of these proteins was selected.

Consensus pattern: [IM]-x-G-x-[GS]-[KRH]-x(4)-[CL]-x-D-G-x(2)-R-x(2)-[RH]-I- x-G

[1] Wei C.-L., Kainuma M., Hershey J.W.B. *J. Biol. Chem.* 270:22788-22794(1995).

182. (eIF-5A) Eukaryotic initiation factor 5A hypusine signature

Eukaryotic initiation factor 5A (eIF-5A) (formerly known as eIF-4D) [1,2] is a small protein whose precise role in the initiation of protein synthesis is not known. It appears to promote the formation of the first peptide bond. eIF-5A seems to be the only eukaryotic protein to contain an hypusine residue. Hypusine is derived from lysine by the post-translational addition of a butylamino group (from spermidine) to the epsilon-amino group of lysine. The hypusine group is essential to the function of eIF-5A. A hypusine-containing protein has been found in archaeobacteria such as *Sulfolobus acidocaldarius* or *Methanococcus jannaschii*; this protein is highly similar to eIF-5A and could play a similar role in protein biosynthesis. The signature developed for eIF-5A is centered around the hypusine residue.

Consensus pattern: [PT]-G-K-H-G-x-A-K [The first K is modified to hypusine]

[1] Park M.H., Wolff E.C., Folk J.E. *Biofactors* 4:95-104(1993).

[2] Schnier J., Schwelberger H.G., Smit-McBride Z., Kang H.A., Hershey J.W.B. *Mol. Cell. Biol.* 11:3105-3114(1991).

183. (eIF-1A) S-100/ICaBP type calcium binding protein signature

S-100 are small dimeric acidic calcium and zinc-binding proteins [1] abundant in the brain. They have two different types of calcium-binding sites: a low affinity one with a special structure and a 'normal' EF-hand type high affinity site. The vitamin-D dependent intestinal calcium-binding proteins (ICaBP or calbindin 9 Kd) also belong to this family of proteins, but it does not form dimers. In the past years the sequences of many new members of this family have been determined (for reviews see [2,3,4]); in most cases the function of these proteins is not yet known, although it is becoming clear that they are involved in cell growth and differentiation, cell cycle regulation and metabolic control. These proteins are: - Calcyclin (Prolactin receptor associated protein (PRA); clatropin; 2a9; 5B10; S100A6). - Calpactin I light chain (p10; p11; 42c; S100A10). - Calgranulin A (cystic fibrosis antigen (CFAg); MIF related protein 8 (MRP- 8); p8; S100A8). - Calgranulin B (MIF related protein 14 (MRP-14); p14; S100A9). - Calgranulin C. - Calgizzarin (S100C). - Placental calcium-binding protein (CAPL) (18a2; pL98; 42a; p9K; MTS1; metastatin; S100A4). - Protein S-100D (S100A5). - Protein S-100E (S100A3). - Protein S-100L (CAN19; S100A2). - Placental protein S-100P (S100E). - Psoriasin (S100A7). - Chemotactic cytokine CP-10 [5]. - Protein MRP-126 [6]. - Trichohyalin [7]. This is a large intermediate filament-associated protein that associates with keratin intermediate filaments (KIF); it contains a S- 100 type domain in its N-terminal extremity. A number of these proteins are known to bind calcium while others are not (p10 for example). Our EF-hand detecting pattern will fail to pick those proteins which have lost their calcium-binding properties. A pattern was developed which unambiguously picks up proteins belonging to this family. This pattern spans the region of the EF-hand high affinity site but makes no assumptions on the calcium-binding properties of this site.

Consensus pattern: [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x- [ES]-[FYVC]-x(2)-[LIVMFS]-[LIVMF]

[1] Baudier J. (In) Calcium and Calcium Binding proteins, Gerday C., Bollis L., Giller R., Eds., pp102-113, Springer Verlag, Berlin, (1988).

[2] Moncrief N.D., Kretsinger R.H., Goodman M. J. Mol. Evol. 30:522-562(1990).

[3] Kligman D., Hilt D.C. Trends Biochem. Sci. 13:437-443(1988).

[4] Schaefer B.W., Wicki R., Engelkamp D., Mattei M.-G., Heizmann C.W. Genomics 25:638-643(1995).

[5] Lackmann M., Cornish C.J., Simpson R.J., Moritz R.L., Geczy C.L. J. Biol. Chem. 267:7499-7504(1992).

[6] Nakano T., Graf T. Oncogene 7:527-534(1992).

[7] Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D., Steinert P.M., J. Biol.

5 Chem. 268:12164-12176(1993).

EF-hand calcium-binding domain

Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand [1 to 5]. This type of

10 domain consists of a twelve residue loop flanked on both side by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12

15 provides two oxygens for liganding Ca (bidentate ligand). Listed below are the proteins which are known to contain EF-hand regions. For each type of protein the total number of EF-hand regions known or supposed to exist is indicated between parenthesis. This number does not include regions which clearly have lost their calcium-binding properties, or the

20 atypical low-affinity site (which spans thirteen residues) found in the S-100/ ICaBP family of proteins [6].

- Aequorin and Renilla luciferin binding protein (LBP) (Ca=3).
- Alpha actinin (Ca=2). - Calbindin (Ca=4).
- Calcineurin B subunit (protein phosphatase 2B regulatory subunit) (Ca=4).
- 25 - Calcium-binding protein from Streptomyces erythraeus (Ca=3?).
- Calcium-binding protein from Schistosoma mansoni (Ca=2?).
- Calcium-binding proteins TCBP-23 and TCBP-25 from Tetrahymena thermophila (Ca=4?).
- Calcium-dependent protein kinases (CDPK) from plants (Ca=4).
- Calcium vector protein from amphoxius (Ca=2).
- 30 - Calcyphosin (thyroid protein p24) (Ca=4?).
- Calmodulin (Ca=4, except in yeast where Ca=3).
- Calpain small and large chains (Ca=2). - Calretinin (Ca=6).
- Calcyclin (prolactin receptor associated protein) (Ca=2).

- Caltractin (centrin) (Ca=2 or 4).
- Cell Division Control protein 31 (gene CDC31) from yeast (Ca=2?).
- Diacylglycerol kinase (EC 2.7.1.107) (DGK) (Ca=2).
- FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) from mammals (Ca=1).
- Fimbrin (plastin) (Ca=2).
- Flagellar calcium-binding protein (1f8) from *Trypanosoma cruzi* (Ca=1 or 2).
- Guanylate cyclase activating protein (GCAP) (Ca=3).
- Inositol phospholipid-specific phospholipase C isozymes gamma-1 and delta-1 (Ca=2) [10].
- Intestinal calcium-binding protein (ICaBPs) (Ca=2).
- MIF related proteins 8 (MRP-8 or CFAG) and 14 (MRP-14) (Ca=2).
- Myosin regulatory light chains (Ca=1).
- Oncomodulin (Ca=2).
- Osteonectin (basement membrane protein BM-40) (SPARC) and proteins that contains an 'osteonectin' domain (QR1, matrix glycoprotein SC1) (see the entry <PDOC00535>) (Ca=1).
- Parvalbumins alpha and beta (Ca=2).
- Placental calcium-binding protein (18a2) (nerve growth factor induced protein 42a) (p9k) (Ca=2).
- Recoverins (visinin, hippocalcin, neurocalcin, S-modulin) (Ca=2 to 3).
- Reticulocalbin (Ca=4).
- S-100 protein, alpha and beta chains (Ca=2).
- Sarcoplasmic calcium-binding protein (SCPs) (Ca=2 to 3).
- Sea urchin proteins Spec 1 (Ca=4), Spec 2 (Ca=4?), Lps-1 (Ca=8).
- Serine/threonine protein phosphatase rdgc (EC 3.1.3.16) from *Drosophila* (Ca=2).
- Sorcin V19 from hamster (Ca=2).
- Spectrin alpha chain (Ca=2).
- Squidulin (optic lobe calcium-binding protein) from squid (Ca=4).
- Troponins C; from skeletal muscle (Ca=4), from cardiac muscle (Ca=3), from arthropods and molluscs (Ca=2).

There has been a number of attempts [7,8] to develop patterns that pick-up EF-hand regions, but these studies were made a few years ago when not so many different families of calcium-binding proteins were known. Therefore a new pattern was developed which takes into account all published sequences. This pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic.

-Consensus pattern: D-x-[DNS]-{ILVFIYW}-[DENSTG]-[DNOGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW]

-Note: positions 1 (X), 3 (Y) and 12 (-Z) are the most conserved.

-Note: the 6th residue in an EF-hand loop is, in most cases a Gly, but the number of exceptions to this 'rule' has gradually increased and therefore the pattern should include all the different residues which have been shown to exist in this position in functional Ca-binding sites.

-Note: the pattern will, in some cases, miss one of the EF-hand regions in some proteins with multiple EF-hand domains.

[1] Kawasaki H., Kretsinger R.H. Protein Prof. 2:305-490(1995).[2] Kretsinger R.H. Cold Spring Harbor Symp. Quant. Biol. 52:499-510(1987).

[3] Moncrief N.D., Kretsinger R.H., Goodman M. J. Mol. Evol. 30:522-562(1990).

[4] Nakayama S., Moncrief N.D., Kretsinger R.H. J. Mol. Evol. 34:416-448(1992).

[5] Heizmann C.W., Hunziker W. Trends Biochem. Sci. 16:98-103(1991).

[6] Kligman D., Hilt D.C. Trends Biochem. Sci. 13:437-443(1988).

[7] Strynadka N.C.J., James M.N.G. Annu. Rev. Biochem. 58:951-98(1989).

[8] Haiech J., Sallantin J. Biochimie 67:555-560(1985).

[9] Chauvaux S., Beguin P., Aubert J.-P., Bhat K.M., Gow L.A., Wood T.M., Bairoch A. Biochem. J. 265:261-265(1990).

[10] Bairoch A., Cox J.A. FEBS Lett. 269:454-456(1990).

184. Enolase signature

Enolase (EC 4.2.1.11) is a glycolytic enzyme that catalyzes the dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate [1]. It is a dimeric enzyme that requires magnesium both for catalysis and stabilizing the dimer. Enolase is probably found in all organisms that metabolize sugars. In vertebrates, there are three different tissue-specific isozymes: alpha present in most tissues, beta in muscles and gamma found only in nervous tissues. Tau-crystallin, one of the major lens proteins in some fish, reptiles and birds, has been shown [2] to be evolutionary related to enolase. As a signature pattern for enolase, the best conserved region was selected, it is located in the C-terminal third of the sequence.-

Consensus pattern: [LIV](3)-K-x-N-Q-l-G-[ST]-[LIV]-[ST]-[DE]-[STA]

[1] Lebioda L., Stec B., Brewer J.M. J. Biol. Chem. 264:3685-3693(1989).

[2] Wistow G., Piattigorsky J. Science 236:1554-1556(1987).

5

185. (F-actin_cap_A) F-actin capping protein alpha subunit signatures

The F-actin capping protein binds in a calcium-independent manner to the fast growing ends of actin filaments (barbed end) thereby blocking the exchange of subunits at these ends.

10 Unlike gelsolin and severin this protein does not sever actin filaments. The F-actin capping protein is a heterodimer composed of two unrelated subunits: alpha and beta. The alpha subunit is a protein of about 268 to 286 amino acid residues whose sequence is well conserved in eukaryotic species [1]. As signature patterns two highly conserved regions in the C-terminal section of the alpha subunit were selected.

15

Consensus pattern: V-H-[FY](2)-E-D-G-N-V

Consensus pattern: F-K-[AE]-L-R-R-x-L-P-

[1] Cooper J.A., Caldwell J.E., Gattermeir D.J., Torres M.A., Amatruda J.F., Casella J.F.

20 Cell Motil. Cytoskeleton 18:204-214(1991).

186. F-box domain

[1] Bai C, Sen P, Hofmann K, Ma L, Goebel M, Harper JW, Elledge SJ, Cell

25 1996;86:263-274. [2] Skowrya D, Craig KL, Tyers M, Elledge SJ, Harper JW, Cell 1997;91:209-219.

187. F-protein

30 Negative factor, (F-Protein) or Nef.

[1] Arold S, Franken P, Strub M-P, Hoh F, Benichou S, Benarous R, Dumas C; Medline: 98035457, The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain

suggests a role for this complex in altered T cell receptor signalling Structure 1997;5:1361-1372.

Nef protein accelerates virulent progression of AIDS by its interaction with cellular proteins involved in signal transduction and host cell activation. Nef has been shown to bind specifically to a subset of the Src kinase family.

Number of members: 1013

188. (FAD_binding_2)

Fumarate reductase / succinate dehydrogenase FAD-binding site

In bacteria two distinct, membrane-bound, enzyme complexes are responsible for the interconversion of fumarate and succinate (EC 1.3.99.1): fumarate reductase (Frd) is used in anaerobic growth, and succinate dehydrogenase (Sdh) is used in aerobic growth. Both complexes consist of two main components: a membrane-extrinsic component composed of a FAD-binding flavoprotein and an iron-sulfur protein; and an hydrophobic component composed of a membrane anchor protein and/or a cytochrome B.

In eukaryotes mitochondrial succinate dehydrogenase (ubiquinone) (EC 1.3.5.1) is an enzyme composed of two subunits: a FAD flavoprotein and and iron-sulfur protein.

The flavoprotein subunit is a protein of about 60 to 70 Kd to which FAD is covalently bound to a histidine residue which is located in the N-terminal section of the protein [1]. The sequence around that histidine is well conserved in Frd and Sdh from various bacterial and eukaryotic species [2] and can be used as a signature pattern.

Consensus patternR-[ST]-H-[ST]-x(2)-A-x-G-G [H is the FAD binding site] Sequences known to belong to this class detected by the pattern ALL.

[1] Blaut M., Whittaker K., Valdovinos A., Ackrell B.A., Gunsalus R.P., Cecchini G. J. Biol. Chem. 264:13599-13604(1989).

[2] Birch-Machin M.A., Farnsworth L., Ackrell B.A., Cochran B., Jackson S., Bindoff L.A., Aitken A., Diamond A.G., Turnbull D.M. J. Biol. Chem. 267:11553-11558(1992).

189. Fatty acid desaturases signatures (FA_desaturase)

Fatty acid desaturases (EC 1.14.99.-) are enzymes that catalyze the insertion of a double bond at the delta position of fatty acids. There seems to be two distinct families of fatty acid desaturases which do not seem to be evolutionary related. Family 1 is composed of: -

Stearoyl-CoA desaturase (SCD) (EC 1.14.99.5) [1]. SCD is a key regulatory enzyme of unsaturated fatty acid biosynthesis. SCD introduces a cis double bond at the delta(9) position of fatty acyl-CoA's such as palmitoleoyl- and oleoyl-CoA. SCD is a membrane-bound enzyme that is thought to function as a part of a multienzyme complex in the endoplasmic reticulum of vertebrates and fungi. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected, this region is rich in histidine residues and in aromatic residues. Family 2 is composed of: - Plants stearoyl-acyl-carrier-protein desaturase (EC 1.14.99.6) [2], these enzymes catalyze the introduction of a double bond at the delta(9) position of stearoyl-ACP to produce oleoyl-ACP. This enzyme is responsible for the conversion of saturated fatty acids to unsaturated fatty acids in the synthesis of vegetable oils. - Cyanobacteria desA [3] an enzyme that can introduce a second cis double bond at the delta(12) position of fatty acid bound to membranes glycerolipids. DesA is involved in chilling tolerance; the phase transition temperature of lipids of cellular membranes being dependent on the degree of unsaturation of fatty acids of the membrane lipids. As a signature pattern for this family a conserved region in the C-terminal part of these enzymes was selected.

Consensus pattern: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y-

Consensus pattern: [ST]-[SA]-x(3)-[OR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]- [DE]-

[1] Kaestner K.H., Ntambi J.M., Kelly T.J. Jr., Lane M.D. J. Biol. Chem. 264:14755-14761(1989).

[2] Shanklin J., Somerville C.R. Proc. Natl. Acad. Sci. U.S.A. 88:2510-2514(1991).

[3] Wada H., Gombos Z., Murata N. Nature 347:200-203(1990).

190. Fructose-1-6-bisphosphatase active site (FBPase)

Fructose-1,6-bisphosphatase (EC 3.1.3.11) (FBPase) [1], a regulatory enzyme in gluconeogenesis, catalyzes the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate. It is involved in many different metabolic pathways and found in most organisms. Sedoheptulose-1,7-bisphosphatase (EC 3.1.3.37) (SBPase) [2] is an enzyme found in plant chloroplast and in photosynthetic bacteria that catalyzes the hydrolysis of sedoheptulose 1,7-bisphosphate to sedoheptulose 7-phosphate, a step in the Calvin's reductive pentose phosphate cycle. It is functionally and structurally related to FBPase. In mammalian FBPase, a lysine residue has been shown to be involved in the catalytic mechanism [3]. The region around this residue is highly conserved and can be used as a signature pattern for FBPase and SBPase. It must be noted that, in some bacterial FBPase sequences, the active site lysine is replaced by an arginine

Consensus pattern: [AG]-[RK]-L-x(1,2)-[LIV]-[FY]-E-x(2)-P-[LIVM]-[GSA] [K/R is the active site residue]-

[1] Benkovic S.J., DeMaine M.M. Adv. Enzymol. 53:45-82(1982).

[2] Raines C.A., Lloyd J.C., Willingham N.M., Potts S., Dyer T.A. Eur. J. Biochem. 205:1053-1059(1992).

[3] Ke H., Thorpe C.M., Seaton B.A., Lipscomb W.N., Marcus F. J. Mol. Biol. 212:513-539(1989).

191. FGGY family of carbohydrate kinases signatures *

It has been shown [1] that four different type of carbohydrate kinases seem to be evolutionary related. These enzymes are: - L-fuculokinase (EC 2.7.1.51) (gene fucK). - Gluconokinase (EC 2.7.1.12) (gene gntK). - Glycerokinase (EC 2.7.1.30) (gene glpK). - Xylulokinase (EC 2.7.1.17) (gene xylB). - L-xylulose kinase (EC 2.7.1.53) (gene lyxK). These enzymes are proteins of from 480 to 520 amino acid residues. As consensus patterns for this family of kinases two conserved regions were selected, one in the central section, the other in the C-terminal section.

Consensus pattern: [MFYGS]-x-[PST]-x(2)-K-[LIVMFYW]-x-W-[LIVMF]-x-[DENQTKR]-[ENQH]-

Consensus pattern: [GSA]-x-[LIVMFYW]-x-G-[LIVM]-x(7,8)-[HDENQ]-[LIVMF]-x(2)-[AS]-[STAIVM]-[LIVMFY]-[DEQ]-

5

[1] Reizer A., Deutscher J., Saier M.H. Jr., Reizer J. Mol. Microbiol. 5:1081-1089(1991).

192. FKBP-type peptidyl-prolyl cis-trans isomerase signatures/profile (FKBP)

- 10 FKBP [1,2,3] is the major high-affinity binding protein, in vertebrates, for the immunosuppressive drug FK506. It exhibits peptidyl-prolyl cis-trans isomerase activity (EC 5.2.1.8) (PPIase or rotamase). PPIase is an enzyme that accelerates protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides [4]. At least three different forms of FKBP are known in mammalian species: - FKBP-12, which is
- 15 cytosolic and inhibited by both FK506 and rapamycin. - FKBP-13, which is membrane associated and inhibited by both FK506 and rapamycin. - FKBP-25, which is preferentially inhibited by rapamycin. These forms of FKBP are evolutionary related and show extensive similarities[5,6,7] with the following proteins: - Fungal FKBP. - Mammalian hsp binding immunophilin (HBI) (also called p59). HBI is a protein which binds to hsp90 and contains
- 20 two FKBP-like domains in its N- terminal section - the first of which seems to be functional. - The C-terminal part of the cell-surface protein mip from Legionella; a protein associated with macrophage infection by an unknown mechanism. - Escherichia coli slyD [8], a protein with a N-terminal FKBP domain followed by an histidine-rich metal-binding domain. - Escherichia coli fkpA. - Escherichia coli fklB (FKBP22). - Escherichia coli slpA. - Bacterial
- 25 trigger factor (Tig). - Streptomyces hygroscopicus and chrysomallus FK506-binding protein. - Chlamydia trachomatis 27 Kd membrane protein. - Neisseria meningitidis strain C114 PPIase. - Probable PPIases from Haemophilus influenzae (HI0754), Methanococcus jannaschii (MJ0278 and MJ0825), Pseudomonas fluorescens and Pseudomonas aeruginosa. Two signature patterns for these proteins were developed. One is based on a conserved region
- 30 in the N-terminus of FKBP, the other is located in the central section. The profile for FKBP spans the complete domain.

[1] Tropschug M., Wachter E., Mayer S., Schoenbrunner E.R., Schmid F.X. Nature 346:674-677(1990).

[3] Siekierka J.J., Widerrecht G., Greulich H., Boulton D., Hung S.H.Y., Cryan J., Hodges P.J., Sigal N.H. J. Biol. Chem. 265:21011-21015(1990).

[5] Trandinh C.C., Pao G.M., Saier M.H. Jr. FASEB J. 6:3410-3420(1992).

[7] Hacker J., Fischer G. Mol. Microbiol. 10:445456(1993).

[8] Wuelfing C., Lomardero J., Plueckthun A. J. Biol. Chem. 269:2895-2901(1994).

The following mammalian proteins are evolutionary related [1]:

- Leukotriene C4 synthase (EC 2.5.1.37) (gene LTC4S), an enzyme that catalyzes the production of LTC4 from LTA4.
- Microsomal glutathione S-transferase II (EC 2.5.1.18) (GST-II) (gene GST2), an enzyme that can also produce LTC4 from LTA4.
- 5-lipoxygenase activating protein (gene FLAP), a protein that seems to be required for the activation of 5-lipoxygenase.

These are proteins of 150 to 160 residues that contain three transmembrane segments. As a signature pattern, a conserved region between the first and second transmembrane domains was selected.

Consensus pattern: G-x(3)-F-E-R-V-[FY]-x-A-[NQ]-x-N-C

[1] Jakobsson P.-J., Mancini J.A., Ford-Hutchinson A.W. J. Biol. Chem. 271:22203-22210(1996).

194. FMN-dependent alpha-hydroxy acid dehydrogenases active site (FMN_dh)

A number of oxidoreductases that act on alpha-hydroxy acids and which are FMN-containing flavoproteins have been shown [1,2,3] to be structurally related; these enzymes are: - Lactate dehydrogenase (EC 1.1.2.3), which consists of a dehydrogenase domain and a heme-binding domain called cytochrome b2 and which catalyzes the conversion of lactate into pyruvate. -

Glycolate oxidase (EC 1.1.3.15) ((S)-2-hydroxy-acid oxidase), a peroxisomal enzyme that catalyzes the conversion of glycolate and oxygen to glyoxylate and hydrogen peroxide. -

Long chain alpha-hydroxy acid oxidase from rat (EC 1.1.3.15), a peroxisomal enzyme. -

Lactate 2-monooxygenase (EC 1.13.12.4) (lactate oxidase) from *Mycobacterium smegmatis*, which catalyzes the conversion of lactate and oxygen to acetate, carbon dioxide and water. -

(S)-mandelate dehydrogenase from *Pseudomonas putida* (gene mdlB), which catalyzes the reduction of (S)-mandelate to benzoylformate. The first step in the reaction mechanism of these enzymes is the abstraction of the proton from the alpha-carbon of the substrate producing a carbanion which can subsequently attach to the N5 atom of FMN. A conserved histidine has been shown [4] to be involved in the removal of the proton. The region around this active site residue is highly conserved and contains an arginine residue which is involved in substrate binding.

Consensus pattern: S-N-H-G-[AG]-R-Q [H is the active site residue] [R is a substrate-binding residue]-

[1] Giegel D.A., Williams C.H. Jr., Massey V. J. Biol. Chem. 265:6626-6632(1990).

[2] Tsou A.Y., Ransom S.C., Gerlt J.A., Buechter D.D., Babbitt P.C., Kenyon G.L.

Biochemistry 29:9856-9862(1990).

[3] Le K.H.D., Lederer F. J. Biol. Chem. 266:20877-20880(1991).

[4] Lindqvist Y., Branden C.-I. J. Biol. Chem. 264:3624-3628(1989).

195. Flavin-binding monooxygenase-like (FMO-like)

This family includes FMO proteins, cyclohexanone monooxygenase

Folylpolyglutamate synthase (EC 6.3.2.17) (FPGS) [1] is the enzyme of folate metabolism that catalyzes ATP-dependent addition of glutamate moieties to tetrahydrofolate.

Its sequence is moderately conserved between prokaryotes (gene *folC*) and eukaryotes. We developed two signature patterns based on the conserved regions which are rich in glycine residues and could play a role in the catalytical activity and/or in substrate binding.

Consensus pattern [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)-[LIVM](2)-x(3)-[GSK] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern[LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2)

Sequences known to belong to this class detected by the pattern ALL.

[1] Shane B., Garrow T., Brenner A., Chen L., Choi Y.J., Hsu J.C., Stover P. Adv. Exp. Med. Biol. 338:629-634(1993).

197. FYVE zinc finger

The FYVE zinc finger is named after four proteins that it has been found in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger has been shown to bind two Zn⁺⁺ ions [1]. The FYVE finger has eight potential zinc coordinating cysteine positions. Many members of this family also include two histidines in a motif R+HHC+XCG, where + represents a charged residue and X any residue. Members were included which do not conserve these histidine residues but are clearly related.

[1] Stenmark H, Aasland R, Toh BH, D'Arrigo A, J Biol Chem 1996;271:24048-24054. [2] Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Stenmark H, Aasland R, Nature 1998;394:432-433.

198. F_actin_cap_B

F-actin capping protein beta subunit signature

The F-actin capping protein binds in a calcium-independent manner to the fast growing ends of actin filaments (barbed end) thereby blocking the exchange of subunits at these ends.

Unlike gelsolin and severin this protein does not sever actin filaments. The F-actin capping protein is a heterodimer composed of two unrelated subunits: alpha and beta.

The beta subunit is a protein of about 280 amino acid residues whose sequence is well conserved in eukaryotic species [1]. As a signature pattern a conserved hexapeptide in the N-terminal section of the beta subunit was selected.

Consensus pattern: C-D-Y-N-R-D Sequences known to belong to this class detected by the pattern ALL.

[1] Amatruda J.F., Cannon J.F., Tatchell K., Hug C., Cooper J.A. Nature 344:352-354(1990).

199. Isopenicillin N synthetase signatures (Fe_Asc_oxidored)

Isopenicillin N synthetase (IPNS) [1,2] is a key enzyme in the biosynthesis of penicillin and cephalosporin. In the presence of oxygen, it removes iron and ascorbate, four hydrogen atoms from L-(alpha-aminoadipyl)-L-cysteinyl-d-valine to form the azetidinone and thiazolidine rings of isopenicillin. IPNS is an enzyme of about 330 amino-acid residues. Two cysteines are conserved in fungal and bacterial IPNS sequences; these may be involved in iron-binding and/or substrate-binding. Cephalosporium acremonium DAOCS/DACS [3] is a bifunctional enzyme involved in cephalosporin biosynthesis. The DAOCS domain, which is structurally related to IPNS, catalyzes the step from penicillin N to deacetoxy-cephalosporin C - used as a substrate by DACS to form deacetylcephalosporin C. Streptomyces clavuligerus possesses a monofunctional DAOCS enzyme (gene cefE) [4] also related to IPNS. Two signature patterns for these enzymes were derived, centered around the conserved cysteine residues.

Consensus pattern: [RK]-x-[STA]-x(2)-S-x-C-Y-[SL]-

Consensus pattern: [LIVM](2)-x-C-G-[STA]-x(2)-[STAG]-x(2)-T-x-[DNG]-

- [1] Martin J.F. Trends Biotechnol. 5:306-308(1987).
 [2] Chen G., Shiffman D., Mevarech M., Aharonowitz Y. Trends Biotechnol. 8:105-111(1990).
 5 [3] Samson S.M., Dotzlar J.E., Slisz M.L., Becker G.W., van Frank R.M., Veal L.E., Yeh W.K., Miller J.R., Queener S.W., Ingolia T.D. Bio/Technology 5:1207-1214(1987).
 [4] Kovacevic S., Weigel B.J., Tobin M.B., Ingolia T.D., Miller J.R. J. Bacteriol. 171:754-760(1989).

200. Fibrillarin signature

Fibrillarin [1] is a component of a nucleolar small nuclear ribonucleoprotein(SnRNP) particle thought to participate in the first step of the processing of pre-rRNA. In mammals, fibrillarin is associated with the U3, U8 and U13small nuclear RNAs [2]. Fibrillarin is an extremely well conserved protein of about 320 amino acid residues. Structurally it consists of three different domains: - An N-terminal domain of about 80 amino acids which is very rich in glycine and contains a number of dimethylated arginine residues (DMA). - A central domain of about 90 residues which resembles that of RNA-binding proteins and contains an octameric sequence similar to the RNP-2 consensus found in such proteins. - A C-terminal alpha-helical domain. A protein evolutionary related to fibrillarin has been found [3] in archaeobacteria such as Methanococcus vannielii or voltae. This protein (geneflpA) is involved in pre-rRNA processing. It lacks the Gly/Arg-rich N-terminal domain. As a signature pattern, a region was selected that starts with and encompasses theRNP-2 like octapeptide sequence.

Consensus pattern: [GST]-[LIVMAP]-V-Y-A-[IV]-E-[FY]-[SA]-x-R-x(2)-R-[DE] -

- [1] Aris J.P., Blobel G. Proc. Natl. Acad. Sci. U.S.A. 88:931-935(1991).
 [2] Bandziulis R.J., Swanson M.S., Dreyfuss G. Genes Dev. 3:431-437(1989).
 30 [3] Agha-Amiri K. J. Bacteriol. 176:2124-2127(1994).

201. Filamin/ABP280 repeat

[1] Fucini P, Renner C, Herberhold C, Noegel AA, Holak TA, Nat Struct Biol 1997;4:223-230.

5 202. Fucosyl transferase

This family of Fucosyltransferases are the enzymes transferring fucose from GDP-Fucose to GlcNAc in an alpha1,3 linkage [1].

[1] Breton C, Oriol R, Imberty A; Glycobiology 1998;8:87-94.

10

203. 2Fe-2S ferredoxins, iron-sulfur binding region signature (fer2A)

Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One of these subgroups are the 2Fe-2S ferredoxins, which are proteins or domains of around one hundred amino acid residues that bind a single 2Fe-2S iron-sulfur cluster. The proteins that are known [2] to belong to this family are listed below. - Ferredoxin from photosynthetic organisms; namely plants and algae where it is located in the chloroplast or cyanelle; and cyanobacteria. - Ferredoxin from archaeobacteria of the Halobacterium genus.

15 - Ferredoxin IV (gene pftA) and V (gene fdxD) from Rhodobacter capsulatus. - Ferredoxin in the toluene degradation operon (gene xylT) and naphthalene degradation operon (gene nahT) of Pseudomonas putida. - Hypothetical Escherichia coli protein yfaE. - The N-terminal domain of the bifunctional ferredoxin/ferredoxin reductase electron transfer component of the benzoate 1,2-dioxygenase complex (gene benC) from Acinetobacter calcoaceticus, the toluene 4-monooxygenase complex (gene tmoF), the toluate 1,2-dioxygenase system (gene xylZ), and the xylene monooxygenase system (gene xylA) from Pseudomonas. - The N-terminal domain of phenol hydroxylase protein p5 (gene dmpP) from Pseudomonas Putida. - The N-terminal domain of methane monooxygenase component C (gene mmoC) from Methylococcus capsulatus. - The C-terminal domain of the vanillate degradation pathway protein vanB in a Pseudomonas species. - The N-terminal domain of bacterial fumarate reductase iron-sulfur protein (gene frdB). - The N-terminal domain of CDP-6-deoxy-3,4-glucoseen reductase (gene ascD) from Yersinia pseudotuberculosis. - The central domain of eukaryotic succinate dehydrogenase (ubiquinone) iron- sulfur protein. - The N-terminal

20

25

30

domain of eukaryotic xanthine dehydrogenase. - The N-terminal domain of eukaryotic aldehyde oxidase. In the 2Fe-2S ferredoxins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region.

Consensus pattern: C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C [The three C's are 2Fe-2S ligands]-

[1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).[2] Harayama S., Polissi A., Rekik M. FEBS Lett. 285:85-88(1991).

Adrenodoxin family, iron-sulfur binding region signature (fer2B)

Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron sulfur cluster(s) and according to sequence similarities. One family of ferredoxins groups together the following proteins that all bind a single 2Fe-2S iron-sulfur cluster: - Adrenodoxin (ADX) (adrenal ferredoxin), a vertebrate mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450_{scc}, which is involved in cholesterol side chain cleavage. - Putidaredoxin (PTX), a *Pseudomonas putida* protein which transfers electrons from putidaredoxin reductase to cytochrome P450_{cam}, which is involved in the oxidation of camphor. - Terpredoxin [2], a *Pseudomonas* protein which transfers electrons from terpredoxin reductase to cytochrome P450_{terp}, which is involved in the oxidation of alpha-terpineol. - Rhodocoxin [3], a *Rhodococcus* protein which transfers electrons from rhodocoxin reductase to cytochrome CYP116 (thcB), which is involved in the degradation of thiocarbamate herbicides. - *Escherichia coli* ferredoxin (gene fdx) [4] whose exact function is not yet known. - *Rhodobacter capsulatus* ferredoxin VI [5], which may transfer electrons to a yet uncharacterized oxygenase. - *Caulobacter crescentus* ferredoxin (gene fdxB) [6]. In these proteins, four cysteine residues bind the iron-sulfur cluster. Three of these cysteines are clustered together in the same region of the protein. Our signature pattern spans that iron-sulfur binding region.

Consensus pattern: C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR] [The three C's are 2Fe-2S ligands]-

[1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).

5 [2] Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S., Carmona C., Witney F., Lorence M.C. J. Biol. Chem. 267:14193-14203(1992).

[3] Nagy I., Schoofs G., Compennolle F., Proost P., Vanderleyden J., De Mot R. J. Bacteriol. 177:676-687(1995).

[4] Ta D.T., Vickery L.E. J. Biol. Chem. 267:11120-11125(1992).

10 [5] Naud I., Vincon M., Garin J., Gaillard J., Forest E., Jouanneau Y. Eur. J. Biochem. 222:933-939(1994).

[6] Amemiya K EMBL/Genbank: X51607.

15 204. 4Fe-4S ferredoxins, iron-sulfur binding region signature (fer4)

Ferredoxins [1] are a group of iron-sulfur proteins which mediate electron transfer in a wide variety of metabolic reactions. Ferredoxins can be divided into several subgroups depending upon the physiological nature of the iron-sulfur cluster(s). One of these subgroups are the 4Fe-4S ferredoxins, which are found in bacteria and which are thus often referred as 'bacterial-type' ferredoxins. The structure of these proteins [2] consists of the duplication of a domain of twenty six amino acid residues; each of these domains contains four cysteine residues that bind to a 4Fe-4S center. A number of proteins have been found [3] that include one or more 4Fe-4S binding domains similar to those of bacterial-type ferredoxins. These proteins are listed below (references are only provided for recently determined sequences). -

25 The iron-sulfur proteins of the succinate dehydrogenase and the fumarate reductase complexes (EC 1.3.99.1). These enzyme complexes, which are components of the tricarboxylic acid cycle, each contain three subunits: a flavoprotein, an iron-sulfur protein, and a b-type cytochrome. The iron-sulfur proteins contain three different iron-sulfur centers: a 2Fe-2S, a 3Fe-3S and a 4Fe-4S. - Escherichia coli anaerobic glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) This enzyme is composed of three subunits: A, B, and C. The C subunit seems to be an iron-sulfur protein with two ferredoxin-like domains in the N-terminal part of the protein. - Escherichia coli anaerobic dimethyl sulfoxide reductase. The B subunit of this enzyme (gene dmsB) is an iron-sulfur protein with four 4Fe-4S ferredoxin-like

domains. - *Escherichia coli* formate hydrogenlyase. Two of the subunits of this oligomeric complex (genes *hycB* and *hycF*) seem to be iron-sulfur proteins that each contain two 4Fe-4S ferredoxin-like domains. - *Methanobacterium formicicum* formate dehydrogenase (EC 1.2.1.2). This enzyme is used by the archaeobacteria to grow on formate. The beta chain of this dimeric enzyme probably binds two 4Fe-4S centers. - *Escherichia coli* formate dehydrogenases N and O (EC 1.2.1.2). The beta chain of these two enzymes (genes *fdnH* and *fdoH*) are iron-sulfur proteins with four 4Fe-4S ferredoxin-like domains. - *Desulfovibrio* periplasmic [Fe] hydrogenase (EC 1.18.99.1). The large chain of this dimeric enzyme binds three 4Fe-4S centers, two of which are located in the ferredoxin-like N-terminal region of the protein. - *Methanobacterium thermoautrophicum* methyl viologen-reducing hydrogenase subunit *mvhB*, which contains six tandemly repeated ferredoxin-like domains and which probably binds twelve 4Fe-4S centers. - *Salmonella typhimurium* anaerobic sulfite reductase (EC 1.8.1.-) [4]. Two of the subunits of this enzyme (genes *asrA* and *asrC*) seem to both bind two 4Fe-4S centers. - A Ferredoxin-like protein (gene *fixX*) from the nitrogen-fixation genes locus of various *Rhizobium* species, and one from the *Nif*-region of *Azotobacter* species. - The 9 Kd polypeptide of chloroplast photosystem I [5] (gene *psaC*). This protein contains two low potential 4Fe-4S centers, referred as the A and B centers. - The chloroplast *frxB* protein which is predicted to carry two 4Fe-4S centers. - An ferredoxin from a primitive eukaryote, the enteric amoeba *Entamoeba histolytica*. - *Escherichia coli* hypothetical protein *yjjW*, a protein with a N-terminal region belonging to the radical activating enzymes family (see <PDOC00834>) and two potential 4Fe-4S centers. The pattern of cysteine residues in the iron-sulfur region is sufficient to detect this class of 4Fe-4S binding proteins.

Consensus pattern: C-x(2)-C-x(2)-C-x(3)-C-[PEG] [The four C's are 4Fe-4S ligands]-

- [1] Meyer J. Trends Ecol. Evol. 3:222-226(1988).
- [2] Otake E., Ooi T. J. Mol. Evol. 26:257-267(1987).
- [3] Beinert H. FASEB J. 4:2483-2492(1990).
- [4] Huang C.J., Barrett E.L. J. Bacteriol. 173:1544-1553(1991).
- [5] Knaff D.B. Trends Biochem. Sci. 13:460-461(1988).

Nitrogenase (EC 1.18.6.1) [1] is the enzyme system responsible for biological nitrogen fixation. Nitrogenase is an oligomeric complex which consists of two components: component 1 which contains the active site for the reduction of nitrogen to ammonia and component 2 (also called the iron protein). Component 2 is a homodimer of a protein (gene *nifH*) which binds a single 4Fe-4S iron sulfur cluster [2]. In the nitrogen fixation process *nifH* is first reduced by a protein such as ferredoxin; the reduced protein then transfers electrons to component 1 with the concomitant consumption of ATP. A number of proteins are known to be evolutionarily related to *nifH*. These proteins are: - Chloroplast encoded *frxC* (or *chlL*) protein [3]. *FrxC* is encoded on the chloroplast genome of some plant species, its exact function is not known, but it could act as an electron carrier in the conversion of protochlorophyllide to chlorophyllide. - *Rhodobacter capsulatus* proteins *bchL* and *bchX* [4]. These proteins are also likely to play a role in chlorophyll synthesis. There are a number of conserved regions in the sequence of these proteins: in the N-terminal section there is an ATP-binding site motif 'A' (P-loop) and in the central section there are two conserved cysteines which have been shown, in *nifH*, to be the ligands of the 4Fe-4S cluster. Two signatures patterns that correspond to the regions around these cysteines were developed.

Consensus pattern: E-x-G-G-P-x(2)-[GA]-x-G-C-[AG]-G [C binds the iron-sulfur center]-

Consensus pattern: D-x-L-G-D-V-V-C-G-G-F-[AG]-x-P [C binds the iron-sulfur center]-

[1] Pau R.N. Trends Biochem. Sci. 14:183-186(1989).

[2] Georgiadis M.M., Komiya H., Chakrabarti P., Woo D., Kornuc J.J., Rees D.C. Science 257:1653-1659(1992).

[3] Fujita Y., Takahashi Y., Kohchi T., Ozeki H., Ohyama K., Matsubara H. Plant Mol. Biol. 13:551-561(1989).

[4] Burke D.H., Alberti M., Hearst J.E. J. Bacteriol. 175:2407-2413(1993).

206. Ferritin iron-binding regions signatures

Ferritin [1,2] is one of the major non-heme iron storage proteins. It consists of a mineral core of hydrated ferric oxide, and a multi-subunit protein shell which englobes the former and assures its solubility in an aqueous environment. In animals the protein is mainly cytoplasmic and there are generally two or more genes that encode for closely related subunits (in

mammals there are two subunits which are known as H(eavy) and L(ight)). In plants ferritin is found in the chloroplast [3]. There are a number of well conserved region in the sequence of ferritins. Two of these regions to develop signature patterns were selected. The first pattern is located in the central part of the sequence of ferritin and it contains three conserved glutamate which are thought to be involved in the binding of iron. The second pattern is located in the C-terminal section, it corresponds to a region which forms a hydrophilic channel through which small molecules and ions can gain access to the central cavity of the molecule; this pattern also includes conserved acidic residues which are potential metal-binding sites.

Consensus pattern: E-x-[KR]-E-x(2)-E-[KR]-[LF]-[LIVMA]-x(2)-Q-N-x-R-x-G-R [The 3 E's are potential iron ligands]-

Consensus pattern: D-x(2)-[LIVMF]-[STAC]-[DH]-F-[LI]-[EN]-x(2)-[FY]-L-x(6)-[LIVM]-[KN] [The second D and the E are potential iron ligands]-

[1] Crichton R.R., Charlotiaux-Wauters M. Eur. J. Biochem. 164:485-506(1987).

[2] Theil E.C. Annu. Rev. Biochem. 56:289-315(1987).

[3] Ragland M., Briat J.-F., Gagnon J., Laulhere J.-P., Massenet O., Theil E.C. J. Biol. Chem. 265:18339-18344(1990).

207. Intermediate filaments signature (filament)

Intermediate filaments (IF) [1,2,3] are proteins which are primordial components of the cytoskeleton and the nuclear envelope. They generally form filamentous structures 8 to 14 nm wide. IF proteins are members of a very large multigene family of proteins which has been subdivided in five major subgroups: - Type I: Acidic cytokeratins. - Type II: Basic cytokeratins. - Type III: Vimentin, desmin, glial fibrillary acidic protein (GFAP), peripherin, and plasticin. - Type IV: Neurofilaments L, H and M, alpha-internexin and nestin. - Type V: Nuclear lamins A, B1, B2 and C. All IF proteins are structurally similar in that they consist of: a central rod domain comprising some 300 to 350 residues which is arranged in coiled-coiled alpha-helices, with at least two short characteristic interruptions; a N-terminal non-helical domain (head) of variable length; and a C-terminal domain (tail) which is also non-helical, and which shows extreme length variation between different IF proteins. While IF proteins are evolutionary and structurally related, they have limited sequence homologies

except in several regions of the rod domain. A conserved region at the C-terminal extremity of the rod domain was used as a sequence pattern for this class of proteins.

Consensus pattern: [IV]-x-[TACI]-Y-[RKH]-x-[LM]-L-[DE]-

[1] Quinlan R., Hutchison C., Lane B. Protein Prof. 2:801-952(1995).

[2] Steiner P.M., Roop D.R. Annu. Rev. Biochem. 57:593-625(1988).

[3] Stewart M. Curr. Opin. Cell Biol. 2:91-100(1990).

208. Flavodoxin signature

Flavodoxins [1,E1] are electron-transfer proteins that function in various electron transport systems. Flavodoxins bind one FMN molecule, which serves as a redox-active prosthetic group. Flavodoxins are functionally interchangeable with ferredoxins. They have been isolated from prokaryotes, cyanobacteria, and some eukaryotic algae. The signature pattern for these proteins is derived from a conserved region in their N-terminal section, this region is involved in the binding of the FMN phosphate group.

Consensus pattern: [LIV]-[LIVFY]-[FY]-x-[ST]-x(2)-[AGC]-x-T-x(3)-A-x(2)-[LIV]-

[1] Wakabayashi S., Kimura K., Matsubara H., Rogers L.J. Biochem. J. 263:981-984(1989).

209. Growth factor and cytokines receptors family signatures (fn3)

A number of receptors for lymphokines, hematopoietic growth factors and growth hormone-related molecules have been found [1 to 5] to share a common binding domain. Receptors known to belong to this family are: - Cytokine receptor common beta chain. This chain is common to the IL-3, IL-5 and GM-CSF receptors. - Cytokine receptor common gamma chain. This chain is common to the IL-2, IL-4, IL-7 and IL-13 receptors. - Ciliary neurotrophic factor receptor (CNTFR). - Erythropoietin receptor (EPOR). - Granulocyte colony-stimulating factor receptor (G-CSFR). - Granulocyte-macrophage colony-stimulating factor receptor alpha chain (GM-CSFR). - Interleukin-2 receptor beta chain (IL2R-beta). - Interleukin-3 receptor alpha chain (IL3R). - Interleukin-4 receptor alpha chain (IL4R). -

Consensus pattern: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W [The two C's are linked by a disulfide bond]-

Consensus pattern: [STGL]-x-W-[SG]-x-W-S-

- [1] Bazan J.F. Biochem. Biophys. Res. Commun. 164:788-795(1989).
- [2] Bazan J.F. Proc. Natl. Acad. Sci. U.S.A. 87:6934-6938(1990).
- [3] Cosman D., Lyman S.D., Idzerda R.L., Beckmann M.P., Park L.S., Goodwin R.G., March C.J. Trends Biochem. Sci. 15:265-270(1990).
- [4] d'Andrea A.D., Fasman G.D., Lodish H.F. Cell 58:1023-1024(1989).
- [5] d'Andrea A.D., Fasman G.D., Lodish H.F. Curr. Opin. Cell Biol. 2:648-651(1990).

210. Phosphoribosylglycinamide formyltransferase active site (formyl transf)

Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) [1] catalyzes the third step in de novo purine biosynthesis, the transfer of a formyl group to 5'-phosphoribosylglycinamide. In higher eukaryotes, GART is part of a multifunctional enzyme polypeptide that catalyzes three of the steps of purine biosynthesis. In bacteria, plants and yeast, GART is a monofunctional protein of about 200 amino-acid residues. In the *Escherichia coli* enzyme, an aspartic acid residue has been shown to be involved in the catalytic mechanism. The region around this active site residue is well conserved in GART from prokaryotic and eukaryotic sources and can be used as a signature pattern. Mammalian

231

formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) [2] is a cytosolic enzyme responsible for the NADP-dependent decarboxylative reduction of 10-formyltetrahydrofolate into tetrahydrofolate. It is a protein of about 900 amino acids consisting of three domains; the N-terminal domain (200 residues) is structurally related to GARTs. *Escherichia coli* methionyl-tRNA formyltransferase (EC 2.1.2.9) (gene *fmt*) [3] is the enzyme responsible for modifying the free amino group of the aminoacyl moiety of methionyl-tRNA (fMet). The central part of *fmt* seems to be evolutionarily related to GART's active site region.

Consensus pattern: G-x-[STM]-[IVT]-x-[FYWVQ]-[VMAT]-x-[DEVN]-x-[LIVMY]-D-x-G-x(2)-[LIVT]-x(6)-[LIVM] [D is the active site residue] -

[1] Inglesse J., Smith J.M., Benkovic S.J. *Biochemistry* 29:6678-6687(1990).

[2] Cook R.J., Lloyd R.S., Wagner C. J. *Biol. Chem.* 266:4965-4973(1991).

[3] Guillon J.-M., Mechulam Y., Schmitter J.-M., Blanquet S., Fayat G. J. *Bacteriol.* 174:4294-4301(1992).

211. G10 protein signatures

A *Xenopus* protein known as G10 [1] has been found to be highly conserved in a wide range of eukaryotic species. The function of G10 is still unknown. G10 is a protein of about 17 to 18 Kd (143 to 157 residues) which is hydrophilic and whose C-terminal half is rich in cysteines and could be involved in metal-binding. As signature patterns, two of these cysteine-rich segments were selected.

Consensus pattern: L-C-C-x-[KR]-C-x(4)-[DE]-x-N-x(4)-C-x-C-R-V-P-

Consensus pattern: C-x-H-C-G-C-[KRH]-G-C-[SA]-

[1] McGrew L.L., Dworkin-Rastl E., Dworkin M.B., Richter J.D. *Genes Dev.* 3:803-815(1989).

212. G-protein alpha subunit

G proteins couple receptors of extracellular signals to intracellular signaling pathways. The G protein alpha subunit binds guanyl nucleotide and is a weak GTPase. Number of members: 195

- 5 [1] Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR, Science 1994;265:1405-1412.
[2] How G proteins work: a continuing story. Coleman DE, Sprang SR, Trends Biochem Sci 1996;21:41-44.

10

213. Glucose-6-phosphate dehydrogenase active site (G6PD)

Glucose-6-phosphate dehydrogenase (EC 1.1.1.49) (G6PD) [1] catalyzes the first step in the pentose pathway, the reduction of glucose-6-phosphate to gluconolactone 6-phosphate. A lysine residue has been identified as are active nucleophile associated with the activity of the enzyme. The sequence around this lysine is totally conserved from bacterial to mammalian G6PD's and can be used as a signature pattern

Consensus pattern: D-H-Y-L-G-K-[EQK] [K is the active site residue]-

- 20 [1] Jeffery J., Persson B., Wood L., Bergman T., Jeffery R., Joernvall H. Eur. J. Biochem. 212:41-49(1993).

214. GATA-type zinc finger domain

- 25 The GATA family of transcription factors are proteins that bind to DNA sites with the consensus sequence (A/T)GATA(A/G), found within the regulatory region of a number of genes. Proteins currently known to belong to this family are: - GATA-1 [1] (also known as Eryf1, GF-1 or NF-E1), which binds to the GATA region of globin genes and other genes expressed in erythroid cells. It is a transcriptional activator which probably serves as a
30 general 'switch' factor for erythroid development. - GATA-2 [2], a transcriptional activator which regulates endothelin-1 gene expression in endothelial cells. - GATA-3 [3], a transcriptional activator which binds to the enhancer of the T-cell receptor alpha and delta genes. - GATA-4 [4], a transcriptional activator expressed in endodermally derived tissues

and heart. - *Drosophila* protein pannier (or DGATAa) (gene *pnr*) which acts as a repressor of the achaete-scute complex (*as-c*). - *Bombyx mori* BCFI [5], which regulates the expression of chorion genes. - *Caenorhabditis elegans* *elt-1* and *elt-2*, transcriptional activators of genes containing the GATA region, including vitellogenin genes [6]. - *Ustilago maydis* *urbs1* [7], a protein involved in the repression of the biosynthesis of siderophores. - Fission yeast protein GAF2. All these transcription factors contain a pair of highly similar 'zinc finger' type domains with the consensus sequence C-x₂-C-x₁₇-C-x₂-C. Some other proteins contain a single zinc finger motif highly related to those of the GATA transcription factors. These proteins are: - *Drosophila* box A-binding factor (ABF) (also known as protein serpent (gene *srp*)) which may function as a transcriptional activator protein and may play a key role in the organogenesis of the fat body. - *Emericella nidulans* *areA* [8], a transcriptional activator which mediates nitrogen metabolite repression. - *Neurospora crassa* *nit-2* [9], a transcriptional activator which turns on the expression of genes coding for enzymes required for the use of a variety of secondary nitrogen sources, during conditions of nitrogen limitation. - *Neurospora crassa* white collar proteins 1 and 2 (WC-1 and WC-2), which control expression of light-regulated genes. - *Saccharomyces cerevisiae* DAL81 (or UGA43), a negative nitrogen regulatory protein. - *Saccharomyces cerevisiae* GLN3, a positive nitrogen regulatory protein. - *Saccharomyces cerevisiae* GAT1. - *Saccharomyces cerevisiae* GZF3.

Consensus pattern: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)-C-N-[AS]-C [The four C's are zinc ligands]

[1] Trainor C.D., Evans T., Felsenfeld G., Boguski M.S. *Nature* 343:92-96(1990).

[2] Lee M.E., Temizer D.T., Clifford J.A., Quertermous T. *J. Biol. Chem.* 266:16188-16192(1991).

[3] Ho L.-C., Vorhees P., Marin N., Oakley B.K., Tsai S.-F., Orkin S.H., Leiden J.M. *EMBO J.* 10:1187-1192(1991).

[4] Spieth J., Shim Y.H., Lea K., Conrad R., Blumenthal T. *Mol. Cell. Biol.* 11:4651-4659(1991).

[5] Drevet J.R., Skeiky Y.A., Iatrou K. *J. Biol. Chem.* 269:10660-10667(1994).

[6] Hawkins M.G., McGhee J.D. *J. Biol. Chem.* 270:14666-14671(1995).

[7] Voisard C.P.O., Wang J., Xu P., Leong S.A., McEvoy J.L. *Mol. Cell. Biol.* 13:7091-7100(1993).

- [8] Arst H.N. Jr., Kudla B., Martinez-Rossi N.M., Caddick M.X., Sibley S., Davies R.W. Trends Genet. 5:291-291(1989).
- [9] Fu Y.-H., Marzluf G.A. Mol. Cell. Biol. 10:1056-1065(1990).

5

215. Glutamine amidotransferases class-I active site (GATase)

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known as glutamine amidotransferase (GATase) (EC 2.4.2.-)

10

[1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I(also known as trpG-type) and class-II (also known as purF-type). Class-I GATase domains have been found in the following enzymes: - The second component of anthranilate synthase (AS) (EC 4.1.3.27) [4]. AS catalyzes the biosynthesis of anthranilate from chorismate and glutamine.

15

AS is generally a dimeric enzyme: the first component can synthesize anthranilate using ammonia rather than glutamine, whereas component II provides the GATase activity. In some bacteria and in fungi the GATase component of AS is part of a multifunctional protein that also catalyzes other steps of the biosynthesis of tryptophan. - The second component of 4-amino-4-deoxychorismate (ADC) synthase (EC 4.1.3. -), a dimeric prokaryotic enzyme that function in the pathway that catalyzes the biosynthesis of para-aminobenzoate (PABA) from chorismate and glutamine. The second component (gene pabA) provides the GATase activity [4]. - CTP synthase (EC 6.3.4.2). CTP synthase catalyzes the final reaction in the biosynthesis of pyrimidine, the ATP-dependent formation of CTP from UTP and glutamine.

25

CTP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the C-terminal section [2]. - GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2). GMP synthase catalyzes the ATP-dependent formation of GMP from xanthosine 5'-phosphate and glutamine. GMP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the N-terminal section [5]. - Glutamine-dependent carbamoyl-phosphate synthase (EC 6.3.5.5) (GD-CPSase); an enzyme involved in both arginine and pyrimidine biosynthesis and which catalyzes the ATP-dependent formation of carbamoyl phosphate from glutamine and carbon dioxide. In bacteria GD-CPSase is composed of two subunits: the large chain (gene carB) provides the CPSase activity, while

30

the small chain (gene *carA*) provides the GATase activity. In yeast the enzyme involved in arginine biosynthesis is also composed of two subunits: CPA1 (GATase), and CPA2 (CPSase). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme (called URA2 in yeast, rudimentary in *Drosophila*, and CAD in mammals). The GATase domain is located at the N-terminal extremity of this polyprotein [6]. - Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3), an enzyme that catalyzes the fourth step in the de novo biosynthesis of purines. In some species of bacteria, FGAM synthase II is composed of two subunits: a small chain (gene *purQ*) which provides the GATase activity and a large chain (gene *purL*) which provides the aminator activity. - The histidine amidotransferase *hisH*, an enzyme that catalyzes the fifth step in the biosynthesis of histidine in prokaryotes. In the second component of AS a cysteine has been shown [7] to be essential for the amidotransferase activity. The sequence around this residue is well conserved in all the above GATase domains and can be used as a signature pattern for class-I GATase.-

Consensus pattern: [PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]-x-[LIVMFA] [C is the active site residue]-

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).

[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).

[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).

[4] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).

[5] Zalkin H., Argos P., Narayana S.V.L., Tiedeman A.A., Smith J.M. J. Biol. Chem. 260:3350-3354(1985).

[6] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[7] Tso J.Y., Hermodson M.A., Zalkin H. J. Biol. Chem. 255:1451-1457(1980).

216. Glutamine amidotransferases class-II active site (GATase_2)

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known as glutamine amidotransferase (GATase) (EC 2.4.2.-) [1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger

polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I(also known as trpG-type) and class-II (also known as purF-type). Class-II GATase domains have been found in the following enzymes: - Amido phosphoribosyltransferase (glutamine phosphoribosylpyrophosphate amidotransferase) (EC 2.4.2.14). An enzyme which catalyzes the first step in purine biosynthesis, the transfer of the ammonia group of glutamine to PRPP to form 5-phosphoribosylamine (gene *purF* in bacteria, *ADE4* in yeast). - Glucosamine--fructose-6-phosphate aminotransferase (EC 2.6.1.16). This enzyme catalyzes a key reaction in amino sugar synthesis, the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine (gene *glmS* in *Escherichia coli*, *nodM* in *Rhizobium*, *GFA1* in yeast) - Asparagine synthetase (glutamine-hydrolyzing) (EC 6.3.5.4). This enzyme is responsible for the synthesis of asparagine from aspartate and glutamine. A cysteine is present at the N-terminal extremity of the mature form of all these enzymes. The cysteine has been shown, in amido phosphoribosyltransferase [4] and in asparagine synthetase [5] to be important for the catalytic mechanism.

Consensus pattern: <x(0,11)-C-[GS]-[IV]-[LIVMFYW]-[AG] [C is the active site residue]-

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).

[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).

[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).

[4] van Heeke G., Schuster M. J. Biol. Chem. 264:5503-5509(1989).

[5] Vollmer S.J., Switzer R.L., Hermodson M.A., Bower S.G., Zalkin H. J. Biol. Chem. 258:10582-10585(1983).

217. GDP dissociation inhibitor (GDI)

[1] Schalk I, Zeng K, Wu SK, Stura EA, Matteson J, Huang M, Tandon A, Wilson IA, Balch WE, Nature 1996;381:42-48.

218. Oxidoreductase family (GFO_IDH_MocA)

This family of enzymes utilise NADP or NAD. This family: is called the GFO/IDH/MOCA family in swiss-prot.

[1] Kingston RL, Scopes RK, Baker EN, Structure 1996;4:1413-1428.

5

219. GHMP kinases putative ATP-binding domain

The following kinases contains, in their N-terminal section, a conserved Gly/Ser-rich region which is probably involved in the binding of ATP [1]. These kinases are listed below. - Galactokinase (EC 2.7.1.6). - Homoserine kinase (EC 2.7.1.39). - Mevalonate kinase (EC 2.7.1.36). - Phosphomevalonate kinase (EC 2.7.4.2). This group of kinases was called 'GHMP' (from the first letter of their substrate)

10

Consensus pattern: [LIVM]-[PK]-x-[GSTA]-x(0,1)-G-L-[GS]-S-S-[GSA]-[GSTAC]-

15

[1] Tsay Y.H., Robinson G.W. Mol. Cell. Biol. 11:620-631(1991).

220. Glucose inhibited division protein A family signatures (GIDA)

Bacterial glucose inhibited division protein A (gene gidA) is a protein of 70Kd whose function is not yet known and whose sequence is highly conserved. It is evolutionary related to yeast hypothetical protein YGL236C, Caenorhabditis elegans hypothetical protein F52H3.2 and a Bacillus subtilis protein called gid (and which is different from B.subtilis gidA). Two highly conserved regions were selected as signature patterns. Both regions are located in the central region of the protein.

25

Consensus pattern: [GS]-[PT]-x-Y-C-P-S-[LIVM]-E-x-K-[LIVM]-x-[KR]-

Consensus pattern: A-G-Q-x-[NT]-G-x(2)-G-Y-x-E-[SAG](3)-[QS]-G-[LIVM](2)-A-G-[LIVMT]-N-A-

30

221. (GLFV_dehydrog)

Glu / Leu / Phe / Val dehydrogenases active site

- Glutamate dehydrogenases (EC 1.4.1.2, EC 1.4.1.3, and EC 1.4.1.4) (GluDH) are enzymes that catalyze the NAD- or NADP-dependent reversible deamination of glutamate into alpha-ketoglutarate [1,2]. GluDH isozymes are generally involved with either ammonia assimilation or glutamate catabolism.
- 5 - Leucine dehydrogenase (EC 1.4.1.9) (LeuDH) is a NAD-dependent enzyme that catalyzes the reversible deamination of leucine and several other aliphatic amino acids to their keto analogues [3].
- Phenylalanine dehydrogenase (EC 1.4.1.20) (PheDH) is a NAD-dependent enzyme that catalyzes the reversible deamidation of L-phenylalanine into phenyl-
10 pyruvate [4].
- Valine dehydrogenase (EC 1.4.1.8) (ValDH) is a NADP-dependent enzyme that catalyzes the reversible deamidation of L-valine into 3-methyl-2-oxobutanoate [5].

15 These dehydrogenases are structurally and functionally related. A conserved lysine residue located in a glycine-rich region has been implicated in the catalytic mechanism. The conservation of the region around this residue allows the derivation of a signature pattern for such type of enzymes.

20 Consensus pattern[LIV]-x(2)-G-G-[SAG]-K-x-[GV]-x(3)-[DNST]-[PL] [K is the active site residue] Sequences known to belong to this class detected by the pattern ALL.

Note all known sequences from this family have Pro in the last position of the pattern with the exception of yeast GluDH which as Leu.

- [1] Britton K.L., Baker P.J., Rice D.W., Stillman T.J. Eur. J. Biochem. 209:851-859(1992).
- [2] Benachenhou-Lahfa N., Forterre P., Labedan B. J. Mol. Evol. 36:335-346(1993).
- [3] Nagata S., Tanizawa K., Esaki N., Sakamoto Y., Ohshima T., Tanaka H., Soda K. Biochemistry 27:9056-9062(1988).
- 30 [4] Takada H., Yoshimura T., Ohshima T., Esaki N., Soda K. J. Biochem. 109:371-376(1991).
- [5] Hutchinson C.R., Tang L. J. Bacteriol. 175:4176-4185(1993).

222. GMC oxidoreductases signatures

The following FAD flavoproteins oxidoreductases have been found [1,2] to be evolutionary related. These enzymes, which are called 'GMC oxidoreductases', are listed below. - Glucose oxidase (EC 1.1.3.4) (GOX) from *Aspergillus niger*. Reaction catalyzed: glucose + oxygen -> delta-gluconolactone + hydrogen peroxide. - Methanol oxidase (EC 1.1.3.13) (MOX) from fungi. Reaction catalyzed: methanol + oxygen -> acetaldehyde + hydrogen peroxide. - Choline dehydrogenase (EC 1.1.99.1) (CHD) from bacteria. Reaction catalyzed: choline + unknown acceptor -> betaine acetaldehyde + reduced acceptor. - Glucose dehydrogenase (GLD) (EC 1.1.99.10) from *Drosophila*. Reaction catalyzed: glucose + unknown acceptor -> delta-gluconolactone + reduced acceptor. - Cholesterol oxidase (CHOD) (EC 1.1.3.6) from *Brevibacterium sterolicum* and *Streptomyces* strain SA-COO. Reaction catalyzed: cholesterol + oxygen -> cholest-4-en-3-one + hydrogen peroxide. - AlkJ [3], an alcohol dehydrogenase from *Pseudomonas oleovorans*, which converts aliphatic medium-chain-length alcohols into aldehydes. This family also includes a lyase: - (R)-mandelonitrile lyase (EC 4.1.2.10) (hydroxynitrile lyase) from plants [4], an enzyme involved in cyanogenesis, the release of hydrogen cyanide from injured tissues. These enzymes are proteins of size ranging from 556 (CHD) to 664 (MOX) amino acid residues which share a number of regions of sequence similarities. One of these regions, located in the N-terminal section, corresponds to the FAD ADP-binding domain. The function of the other conserved domains is not yet known; two of these domains were selected as signature patterns. The first one is located in the N-terminal section of these enzymes, about 50 residues after the ADP-binding domain, while the second one is located in the central section.

Consensus pattern: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-[PAG]-x(5)-[DNESH]-
 Consensus pattern: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G-

[1] Cavener D.R. J. Mol. Biol. 223:811-814(1992).

[2] Henikoff S., Henikoff J.G. Genomics 19:97-107(1994).

[3] van Beilen J.B., Eggink G., Enequist H., Bos R., Witholt B. Mol. Microbiol. 6:3121-3136(1992).

[4] Cheng I.P., Poulton J.E. Plant Cell Physiol. 34:1139-1143(1993).

223. (GMP_synt_C)

Glutamine amidotransferases class-I active site

A large group of biosynthetic enzymes are able to catalyze the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. This catalytic activity is known as glutamine amidotransferase (GATase) (EC 2.4.2.-) [1]. The GATase domain exists either as a separate polypeptidic subunit or as part of a larger polypeptide fused in different ways to a synthase domain. On the basis of sequence similarities two classes of GATase domains have been identified [2,3]: class-I (also known as trpG-type) and class-II (also known as purF-type). Class-I GATase domains have been found in the following enzymes:

- The second component of anthranilate synthase (AS) (EC 4.1.3.27) [4]. AS catalyzes the biosynthesis of anthranilate from chorismate and glutamine. AS is generally a dimeric enzyme: the first component can synthesize anthranilate using ammonia rather than glutamine, whereas component II provides the GATase activity. In some bacteria and in fungi the GATase component of AS is part of a multifunctional protein that also catalyzes other steps of the biosynthesis of tryptophan.
- The second component of 4-amino-4-deoxychorismate (ADC) synthase (EC 4.1.3. -), a dimeric prokaryotic enzyme that function in the pathway that catalyzes the biosynthesis of para-aminobenzoate (PABA) from chorismate and glutamine. The second component (gene pabA) provides the GATase activity [4].
- CTP synthase (EC 6.3.4.2). CTP synthase catalyzes the final reaction in the biosynthesis of pyrimidine, the ATP-dependent formation of CTP from UTP and glutamine. CTP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the C-terminal section [2].
- GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2). GMP synthase catalyzes the ATP-dependent formation of GMP from xanthosine 5'-phosphate and glutamine. GMP synthase is a single chain enzyme that contains two distinct domains; the GATase domain is in the N-terminal section [5].

- Glutamine-dependent carbamoyl-phosphate synthase (EC 6.3.5.5) (GD-CPSase); an enzyme involved in both arginine and pyrimidine biosynthesis and which catalyzes the ATP-dependent formation of carbamoyl phosphate from glutamine and carbon dioxide. In bacteria GD-CPSase is composed of two subunits: the large chain (gene *carB*) provides the CPSase activity, while the small chain (gene *carA*) provides the GATase activity. In yeast the enzyme involved in arginine biosynthesis is also composed of two subunits: CPA1 (GATase), and CPA2 (CPSase). In most eukaryotes, the first three steps of pyrimidine biosynthesis are catalyzed by a large multifunctional enzyme (called URA2 in yeast, rudimentary in *Drosophila*, and CAD in mammals). The GATase domain is located at the N-terminal extremity of this polyprotein [6].

- Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3), an enzyme that catalyzes the fourth step in the de novo biosynthesis of purines. In some species of bacteria, FGAM synthase II is composed of two subunits: a small chain (gene *purQ*) which provides the GATase activity and a large chain (gene *purL*) which provides the aminator activity.

- The histidine amidotransferase *hisH*, an enzyme that catalyzes the fifth step in the biosynthesis of histidine in prokaryotes.

In the second component of AS a cysteine has been shown [7] to be essential for the amidotransferase activity. The sequence around this residue is well conserved in all the above GATase domains and can be used as a signature pattern for class-I GATase.

Consensus pattern[PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]- x-[LIVMFA] [C is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for 6 sequences.

Note: in the first position of the pattern Pro is found in all cases except in the slime mold GD-CPSase where it is replaced by Ala.

[1] Buchanan J.M. Adv. Enzymol. 39:91-183(1973).

[2] Weng M., Zalkin H. J. Bacteriol. 169:3023-3028(1987).

[3] Nyunoya H., Lusty C.J. J. Biol. Chem. 259:9790-9798(1984).

[4] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).

[5] Zalkin H., Argos P., Narayana S.V.L., Tiedeman A.A., Smith J.M. J. Biol. Chem. 260:3350-3354(1985).

[6] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

5 [7] Tso J.Y., Hermodson M.A., Zalkin H. J. Biol. Chem. 255:1451-1457(1980).

224. Glutathione peroxidases signatures (GSHPx)

10 Glutathione peroxidase (EC 1.11.1.9) (GSHPx) [1,2] is an enzyme that catalyzes the reduction of hydroxyperoxides by glutathione. Its main function is to protect against the damaging effect of endogenously formed hydroxyperoxides. In higher vertebrates at least four forms of GSHPx are known to exist: a ubiquitous cytosolic form (GSHPx-1), a gastrointestinal cytosolic for (GSHPx-GI) [3], a plasma secreted form (GSHPx-P) [4], and a epididymal secretory form (GSHPx-EP). In addition to these characterized forms, the sequence of a protein of unknown function [5] has been shown to be evolutionary related to those of GSHPx's. In filarial nematode parasites such as *Brugia pahangi* the major soluble cuticular protein, known as gp29, is a secreted GSHPx which could provide a mechanism of resistance to the immune reaction of the mammalian host by neutralizing the products of the oxidative burst of leukocytes [6]. *Escherichia coli* protein btuE, a periplasmic protein involved in the transport of vitamin B12, is also evolutionary related to GSHPx's; the significance of this relationship is not yet clear. Selenium, in the form of selenocysteine [7] is part of the catalytic site of GSHPx. The sequence around the selenocysteine residue is moderately well conserved in GSHPx's and the related proteins and can be used as a signature pattern. As a second signature for this family of proteins a highly conserved octapeptide located in the central section of these proteins was selected.

15
20
25

Consensus pattern: [GN]-[RKHNFCY]-x-[LIVMFC]-[LIVMF](2)-x-N-[VT]-x-[STC]-x-C-[GA]-x-T [C is the active site selenocysteine residue]

Consensus pattern: [LIV]-[AGD]-F-P-[CS]-[NG]-Q-

30 [1] Mannervik B. Meth. Enzymol. 113:490-495(1985).

[2] Mullenbach G.T., Tabrizi A., Irvine B.D., Bell G.I., Tainer J.A., Hallewell R.A. Protein Eng. 2:239-246(1988).

- [3] Chu F.F., Doroshow J.H., Esworthy R.S. J. Biol. Chem. 268:2571-2576(1993).
 [4] Takahashi K., Akasaka M., Yamamoto Y., Kobayashi C., Mizoguchi J., Koyama J. J. Biochem. 108:145-148(1990).
 [5] Dunn D.K., Howells D.D., Richardson J., Goldfarb P.S. Nucleic Acids Res. 17:6390-6390(1989).
 [6] Cookson E., Blaxter M.L., Selkirk M.E. Proc. Natl. Acad. Sci. U.S.A. 89:5837-5841(1992).
 [7] Stadtman T.C. Annu. Rev. Biochem. 59:111-127(1990).

225. (GST)

Glutathione S-transferases

Function: conjugation of reduced glutathione to a variety of targets. Also included in the alignment, but are not GSTs S-crystallins from squid. Similarity to GST was previously noted. Eukaryotic elongation factors 1-gamma. Not known to have GST activity; similarity not previously recognized. Supported by HMM and manual alignment inspection. HSP26 family of stress-related proteins. including auxin-regulated proteins in plants and stringent starvation proteins in E. coli. Not known to have GST activity. Similarity not previously recognized. Supported by HMM and manual alignment inspection. Alignment spans entire protein.

226. GTP1/OBG family signature

A widespread family of GTP-binding proteins has been recently characterized [1,2]. This family currently includes: - Mouse and Xenopus protein DRG. - Human protein DRG2. - Drosophila protein 128up. - Fission yeast protein gtp1. - A Halobacterium cutirubrum hypothetical protein in a ribosomal protein gene cluster. - Bacillus subtilis protein obg. Obg has been experimentally shown to bind GTP. - Escherichia coli hypothetical protein yhbZ. - Haemophilus influenzae hypothetical protein HI0877. - Mycoplasma genitalium hypothetical protein MG384. - Yeast hypothetical protein YAL036c (FUN11). - Yeast hypothetical protein YGR173w. - Caenorhabditis elegans hypothetical protein C02F5.3. The function of the proteins that belong to this family is not yet known. They are polypeptides of about 40 to

48 Kd which contain the five small sequence elements characteristic of GTP-binding proteins [3]. As a signature pattern the region that correspond to the ATP/GTP B motif (also called G-3 in GTP-binding proteins) was selected.

5 Consensus pattern: D-[LIVM]-P-G-[LIVM](2)-[DEY]-[GN]-A-x(2)-G-x-G -

[1] Sazuka T., Tomooka Y., Ikawa Y., Noda M., Kumar S. Biochem. Biophys. Res. Commun. 189:363-370(1992).

[2] Hudson J.D., Young P.G. Gene 125:191-193(1993).

10 [3] Bourne H.R., Sanders D.A., McCormick F. Nature 349:117-127(1991).

227. (GTP_EFTU1)

ATP/GTP-binding site motif A (P-loop)

15 From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is

20 generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5]. There are numerous ATP- or GTP-binding proteins in which the P-loop is found. Listed below are a number of protein families for which the relevance of the presence of such motif has been noted: - ATP synthase alpha and beta subunits (see <PDOC00137>). - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>). - Dynamins and dynamin-like

25 proteins (see <PDOC00362>). - Guanylate kinase (see <PDOC00670>). - Thymidine kinase (see <PDOC00524>). - Thymidylate kinase (see <PDOC01034>). - Shikimate kinase (see <PDOC00868>). - Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>). - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>). - DNA and RNA helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). - Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).

30 - Nuclear protein ran (see <PDOC00859>). - ADP-ribosylation factors family (see <PDOC00781>). - Bacterial dnaA protein (see <PDOC00771>). - Bacterial recA protein (see <PDOC00131>). - Bacterial recF protein (see <PDOC00539>). - Guanine nucleotide-binding

proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family (See <PDOC00388>). - Bacterial type II secretion system protein E (see <PDOC00567>). Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

-Consensus pattern: [AG]-x(4)-G-K-[ST]-

- [1] Walker J.E., Saraste M., Runswick M.J., Gay N.J. EMBO J. 1:945-951(1982).
- [2] Moller W., Amons R. FEBS Lett. 186:1-7(1985).
- [3] Fry D.C., Kuby S.A., Mildvan A.S. Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).
- [4] Dever T.E., Glynnias M.J., Merrick W.C. Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).
- [5] Saraste M., Sibbald P.R., Wittinghofer A. Trends Biochem. Sci. 15:430-434(1990).
- [6] Koonin E.V. J. Mol. Biol. 229:1165-1174(1993).
- [7] Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P. J. Bioenerg. Biomembr. 22:571-592(1990).
- [8] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).
- [9] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).
- [10] Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. Nucleic Acids Res. 17:4713-4730(1989).

GTP-binding elongation factors signature (GTP_EFTU2)

Elongation factors [1,2] are proteins catalyzing the elongation of peptide chains in protein biosynthesis. In both prokaryotes and eukaryotes, there are three distinct types of elongation factors, as described in the following table: -----

-----	Eukaryotes	Prokaryotes	Function	-----
-----	EF-1alpha	EF-Tu	Binds GTP and an aminoacyl-tRNA; delivers the latter to	-----

the A site of ribosomes. EF-1beta EF-Ts Interacts with EF-1a/EF-Tu to displace GDP and thus allows the regeneration of GTP-EF-1a. EF-2 EF-G Binds GTP and peptidyl-tRNA and translocates the latter from the A site to the P site. -----

-----The GTP-binding elongation factor family also includes the following

5 proteins: - Eukaryotic peptide chain release factor GTP-binding subunits [3]. These proteins interact with release factors that bind to ribosomes that have encountered a stop codon at their decoding site and help them to induce release of the nascent polypeptide. The yeast protein was known as SUP2 (and also as SUP35, SUF12 or GST1) and the human homolog as
10 GST1-Hs. - Prokaryotic peptide chain release factor 3 (RF-3) (gene prfC). RF-3 is a class-II RF, a GTP-binding protein that interacts with class I RFs (see <PDOC00607>) and enhance their activity [4]. - Prokaryotic GTP-binding protein lepA and its homolog in yeast (gene GUF1) and in *Caenorhabditis elegans* (ZK1236.1). - Yeast HBS1 [5]. - Rat statin S1 [6], a protein of unknown function which is highly similar to EF-1alpha. - Prokaryotic
15 selenocysteine-specific elongation factor selB [7], which seems to replace EF-Tu for the insertion of selenocysteine directed by the UGA codon. - The tetracycline resistance proteins tetM/tetO [8,9] from various bacteria such as *Campylobacter jejuni*, *Enterococcus faecalis*, *Streptococcus mutans* and *Ureaplasma urealyticum*. Tetracycline binds to the prokaryotic ribosomal 30S subunit and inhibits binding of aminoacyl-tRNAs. These proteins abolish the inhibitory effect of tetracycline on protein synthesis. - *Rhizobium* nodulation protein nodQ
20 [10]. - *Escherichia coli* hypothetical protein yihK [11]. In EF-1-alpha, a specific region has been shown [12] to be involved in a conformational change mediated by the hydrolysis of GTP to GDP. This region is conserved in both EF-1alpha/EF-Tu as well as EF-2/EF-G and thus seems typical for GTP-dependent proteins which bind non-initiator tRNAs to the ribosome. The pattern developed for this family of proteins include that conserved region.

25 Consensus pattern: D-[KRSTGANQFYW]-x(3)-E-[KRAQ]-x-[RKQD]-[GC]-[IVMK]-[ST]-[IV]-x(2)-[GSTACKRNQ]-

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

[2] Moldave K. Annu. Rev. Biochem. 54:1109-1149(1985).

[3] Stansfield I., Jones K.M., Kushnirov V.V., Dagkesamanskaya A.R., Poznyakovski A.I., Paushkin S.V., Nierras C.R., Cox B.S., Ter-Avanesyan M.D., Tuite M.F. *EMBO J.* 14:4365-4373(1995).

[4] Grentzmann G., Brechemier-Baey D., Heurgue-Hamard V., Buckingham R.H. *J. Biol. Chem.* 270:10595-10600(1995).

[5] Nelson R.J., Ziegelhoffer T., Nicolet C., Werner-Washburne M., Craig E.A. *Cell* 71:97-105(1992).

[6] Ann D.K., Moutsatsos I.K., Nakamura T., Lin H.H., Mao P.-L., Lee M.-J., Chin S., Liem R.K.H., Wang E. *J. Biol. Chem.* 266:10429-10437(1991).

[7] Forchhammer K., Leinfelder W., Bock A. *Nature* 342:453-456(1989).

[8] Manavathu E.K., Hiratsuka K., Taylor D.E. *Gene* 62:17-26(1988).

[9] Leblanc D.J., Lee L.N., Titmas B.M., Smith C.J., Tenover F.C. *J. Bacteriol.* 170:3618-3626(1988).

[10] Cervantes E., Sharma S.B., Maillet F., Vasse J., Truchet G., Rosenberg C. *Mol. Microbiol.* 3:745-755(1989).

[11] Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R. *Nucleic Acids Res.* 21:3391-3398(1993).

[12] Moller W., Schipper A., Amons R. *Biochimie* 69:983-989(1987).

228. GTP cyclohydrolase II.

GTP cyclohydrolase II catalyses the first committed step in the biosynthesis of riboflavin.

[1] Richter G, Ritz H, Katzenmeier G, Volk R, Kohnle A, Lottspeich F, Allendorf D, Bacher A, *J Bacteriol* 1993;175:4045-4051.

229. Galactose-1-phosphate uridyl transferase signatures (GalP_UDP_transf)

Galactose-1-phosphate uridyl transferase (EC 2.7.7.10) (galT) catalyzes the transfer of an uridyldiphosphate group on galactose (or glucose) 1-phosphate. During the reaction, the uridyl moiety links to a histidine residue. In the *Escherichia coli* enzyme, it has been shown [1] that two histidine residues separated by a single proline residue are essential for enzyme activity. On the basis of sequence similarities, two apparently unrelated families seem to

exist. Class-I enzymes are found in eukaryotes as well as some bacteria such as *Escherichia coli* or *Streptomyces lividans*, while class-II enzymes have been found so far only in bacteria such as *Bacillus subtilis* or *Lactobacillus helveticus* [2]. Signature patterns for both families were developed. For class-I enzymes the signature is based on the active site residues. For class-II enzymes a region which also includes two conserved histidines was chosen.

Consensus pattern: F-E-N-[RK]-G-x(3)-G-x(4)-H-P-H-x-Q [The two H's are the active site residues]-

Consensus pattern: D-L-P-I-V-G-G-[ST]-[LIVM](2)-[SA]-H-[DEN]-H-[FY]-Q-G-G -

Note: class-I enzymes are structurally related to the HIT family of proteins (see <PDOC00694

[1] Reichardt J.K.V., Berg P. *Nucleic Acids Res.* 16:9017-9026(1988).

[2] Mollet B., Pilloud N. J. *Bacteriol.* 173:4464-4473(1991).

230. Gamma-thionins family signature

The following small plant proteins are evolutionary related:

- Gamma-thionins from wheat endosperm (gamma-purothionins) and barley (gamma- hordothionins) which are toxic to animal cells and inhibit protein synthesis in cell free systems [1].
- A flower-specific thionin (FST) from tobacco [2].
- Antifungal proteins (AFP) from the seeds of Brassicaceae species such as radish, mustard, turnip and *Arabidopsis thaliana* [3].
- Inhibitors of insect alpha-amylases from sorghum [4].
- Probable protease inhibitor P322 from potato.
- A germination-related protein from cowpea [5].
- Anther-specific protein SF18 from sunflower [6]. SF18 is a protein that contains a gamma-thionin domain at its N-terminus and a proline-rich C- terminal domain.
- Soybean sulfur-rich protein SE60 [7].
- *Vicia faba* antibacterial peptides fabatin-1 and -2.

In their mature form, these proteins generally consist of about 45 to 50 amino-acid residues. As shown in the following schematic representation, these peptides contain eight conserved cysteines involved in disulfide bonds.

+-----+ | +-----+ ||||

5 xxCxxxxxxxxxCxxxxCxxxCxxxxxxxxxCxxxxxCxCxxxC *****|**||

+---|-----+ | +-----+

'C': conserved cysteine involved in a disulfide bond.

'*': position of the pattern.

10 Consensus pattern: [KRG]-x-C-x(3)-[SV]-x(2)-[FYWH]-x-[GF]-x-C-x(5)-C-x(3)-C [The four C's are involved in disulfide bonds]-

[1] Bruix M., Jimenez M.A., Santoro J., Gonzalez C., Colilla F.J., Mendez E., Rico M. Biochemistry 32:715-724(1993).

15 [2] Gu Q., Kawata E.E., Morse M.-J., Wu H.-M., Cheung A.Y. Mol. Gen. Genet. 234:89-96(1992).

[3] Terras F.R.G., Torreken S., van Leuven F., Osborn R.W., Vanderleyden J., Cammue B.P.A., Broekaert W.F. FEBS Lett. 316:233-240(1993).

[4] Bloch C. Jr., Richardson M. FEBS Lett. 279:101-104(1991).

20 [5] Ishibashi N., Yamauchi D., Miniamikawa T. Plant Mol. Biol. 15:59-64(1990).

[7] Choi Y., Choi Y.D., Lee J.S. Plant Physiol. 101:699-700(1993).

231. Gelsolin. Gelsolin repeat. Number of members: 170

25

[1] Medline: 97433077. The crystal structure of plasma gelsolin: implications for actin severing, capping, and nucleation. Burtinck LD, Koepf EK, Grimes J, Jones EY, Stuart DI, McLaughlin PJ, Robinson RC; Cell 1997;90:661-670.

30

232. Germin family signature

Germins [1] are a family of homopentameric cereal glycoproteins expressed during germination which may play a role in altering the properties of cell walls during germinative

5

10

15

20

25

30

GluTR is a protein of about 50 Kd (467 to 550 residues) which contains a few conserved region. The best conserved region is located in positions 99 to 122 in the sequence of known GluTR. This region seems important for the activity of the enzyme. We have developed a signature pattern from that conserved region.

Consensus patternH-[LIVM]-x(2)-[LIVM]-[GSTAC](3)-[LIVM]-[DEQ]-S-[LIVMA]-
[LIVM](2)-[GF]-E-x-[EQR]-[IV]-[LIT]-[STAG]-Q-[LIVM]-[KR] Sequences known to
belong to this class detected by the pattern ALL.

[1] Jahn D., Verkamp E., Soell D. Trends Biochem. Sci. 17:215-218(1992).

234. (Glycoprotease)

Glycoprotease family signature (aka Peptidase_M22)

Glycoprotease (GCP) (EC 3.4.24.57) [1], or o-sialoglycoprotein endopeptidase, is a
metalloprotease secreted by *Pasteurella haemolytica* which specifically cleaves O-
sialoglycoproteins such as glycophorin A. The sequence of GCP is highly similar to the
following uncharacterized proteins:

- *Escherichia coli* hypothetical protein ygiD (ORF-X).
- *Bacillus subtilis* hypothetical protein ydiE.
- *Mycobacterium leprae* hypothetical protein U229E.
- *Mycobacterium tuberculosis* hypothetical protein MtCY78.10.
- *Synechocystis* strain PCC 6803 hypothetical protein slr0807.
- *Methanococcus jannaschii* hypothetical protein MJ1130.
- *Haloarcula marismortui* hypothetical protein in HSH 3' region.
- Yeast hypothetical protein YKR038c.
- Yeast hypothetical protein QRI7.

One of the conserved regions contains two conserved histidines. It is possible that this region
is involved in coordinating a metal ion such as zinc.

Consensus pattern[KR]-[GSAT]-x(4)-[FYWLH]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H- x(2)-
[AG]-H-[LIVM] Sequences known to belong to this class detected by the pattern ALL.

Note: these proteins belong to family M22 in the classification of peptidases [2,E1].

□

[1] Abdullah K.M., Lo R.Y.C., Mellors A. J. Bacteriol. 173:5597-5603(1991).

5 □

[2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

235. (Glucosamine_iso)

10 Glucosamine/galactosamine-6-phosphate isomerases signature

Glucosamine-6-phosphate isomerase (EC 5.3.1.10) (or Glc-6-P deaminase) is the enzyme responsible for the conversion of glucosamine 6-phosphate into fructose6 phosphate [1]. It is the last specific step in the pathway for N-acetylglucosamine (GlcNAC) utilization in bacteria such as *Escherichia coli* (gene nagB) or in fungi such as *Candida albicans* (gene NAG1).Glc-6-P isomerase is evolutionary related to: - A putative *Escherichia coli* galactosamine-6-phosphate isomerase (gene agal) [2]. - *Escherichia coli* hypothetical protein yieK. - *Bacillus subtilis* hypothetical protein ybft. As a signature pattern a conserved region located in the central part of these enzymes was selected. This region contains a conserved histidine which has been shown [1], in nagB, to be important for the pyranose ring-opening step of the catalytic mechanism

Consensus pattern: [LIVM]-x(3)-G-x-[LIT]-x-[LIV]-x-[LIVM]-x-G-[LIVM]-G-x- [DEN]-G-H-

25 [1] Oliva G., Fontes M.R.M., Garratt R.C., Altamirano M.M., Calcagno M.L., Horjaes E. Structure 3:1323-1332(1995).

[2] Reizer J., Ramseier T.M., Reizer A., Charbit A., Saier M.H. Jr. Microbiology 142:231-250(1996).

30

236. Pneumovirus attachment glycoprotein G (glycoprotein G)

This family includes attachment proteins from respiratory syncytial virus. Glycoprotein G has not been shown to have any neuraminidase or hemagglutinin activity (Swiss-Prot). The

253

amino terminus is thought to be cytoplasmic, and the carboxyl terminus extracellular. The extracellular region contains four completely conserved cysteine residues.

[1] Johnson PR, Spriggs MK, Olmsted RA, Collins PL, Proc Natl Acad Sci U S A 1987;84:5625-5629.

5

237. Glycosyl transferases group 1

Mutations in this domain of [Swiss:P37287](#) lead to disease (Paroxysmal Nocturnal haemoglobinuria). Members of this family transfer activated sugars to a variety of substrates, including glycogen, Fructose-6-phosphate and lipopolysaccharides. Members of this family transfer UDP, ADP, GDP or CMP linked sugars. The eukaryotic glycogen synthases may be distant members of this family.

10

15

238. Glycosyl transferases (Glycos_transf_2)

Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl-galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids.

20

239. (Glucos_transf_3)

Thymidine and pyrimidine-nucleoside phosphorylases signature

25

Thymidine phosphorylase (EC 2.4.2.4) catalyzes the reversible phosphorolysis of thymidine, deoxyuridine and their analogues to their respective bases and 2-deoxyribose 1-phosphate. This enzyme regulates the availability of thymidine and is therefore essential to nucleic acid metabolism.

30

In *Escherichia coli* (gene *deoA*), the enzyme is a dimer of identical subunits of about 48 Kd [1]. In humans it was first identified as platelet-derived endothelial cell growth factor (PD-ECGF) [E1] before being recognized [2] as thymidine phosphorylase.

Bacterial pyrimidine-nucleoside phosphorylase (EC 2.4.2.2) (gene pdp) [3] is an enzyme evolutionary and structurally related to thymidine phosphorylase.

A well conserved region of 19 residues located in the N-terminal part of these proteins signature pattern for these enzymes was selected.

Consensus pattern S-[GS]-R-[GA]-[LIV]-x(2)-[TA]-[GA]-G-T-x-D-x-[LIV]-E Sequences known to belong to this class detected by the pattern ALL.

[1] Walter M.R., Cook W.J., Cole L.B., Short S.A., Koszalka G.W., Krenitsky T.A., Ealick S.E. J. Biol. Chem. 265:14016-14022(1990).

[2] Furukawa T., Yoshimura A., Sumizawa T., Haraguchi M., Akiyama S.-I., Fukui K., Yamada Y. Nature 356:668-668(1992).

[3] Saxild H.H., Andersen L.N., Hammer K. J. Bacteriol. 178:424-434(1996).

240. Glycosyl_transf_4. Glycosyl transferase. Number of members: 44.

[1] Medline: 95252686. A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases. Lehrman MA; Glycobiology 1994;4:768-771.

241. Glycosyl hydrolases family 15. 21 members.

242. Glycosyl hydrolases family 16 signature

It has been shown [1] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities: - Bacterial beta-1,3-1,4-glucanases, or lichenases, (EC 3.2.1.73) mainly from *Bacillus* but also from *Clostridium thermocellum* (gene *licB*), *Fibrobacter succinogenes* and *Rhodothermus marinus* (gene *bglA*). - *Bacillus circulans* beta-1,3-glucanase A1 (EC 3.2.1.39) (gene *glcA*). - Laminarinase (EC 3.2.1.6) from *Clostridium thermocellum* (gene *lam1*). - *Streptomyces coelicolor* agarase (EC 3.2.1.81) (gene *dagA*). - *Alteromonas carrageenovora* kappa-carrageenase (EC 3.2.1.83) (gene

cgkA). Two closely clustered conserved glutamates have been shown [2] to be involved in the catalytic activity of *Bacillus licheniformis* lichenase. The region was used that contains these residues as a signature pattern.

- 5 Consensus pattern: E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA] [The two E's are active site residues]-

[1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Juncosa M., Pons J., Dot T., Querol E., Planas A. *J. Biol. Chem.* 269:14530-

10 14535(1994).

243. Glycosyl hydrolases family 17 signature

- 15 It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities: - Glucan endo-1,3-beta-glucosidases (EC 3.2.1.39) (endo-(1->3)-beta- glucanase) from various plants. This enzyme may be involved in the defense of plants against pathogens through its ability to degrade fungal cell wall polysaccharides. - Glucan 1,3-beta-glucosidase (EC 3.2.1.58) (exo-(1->3)-beta-glucanase) from yeast (gene BGL2). This enzyme may play a role in cell expansion during growth, in cell-cell fusion during mating, and in spore release during sporulation. - Lichenases (EC 20 3.2.1.73) (endo-(1->3,1->4)-beta-glucanase) from various plants. The best conserved region in the sequence of these enzymes is located in their central section. This region contains a conserved tryptophan residue which could be involved in the interaction with the glucan substrates [2] and it also contains a conserved glutamate which has been shown [3] to act as
- 25 the nucleophile in the catalytic mechanism. this region was used as a signature pattern.

Consensus pattern: [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ] [E is an active site residue]-

30 [1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Ori N., Sessa G., Lotan T., Himmelhoch S., Fluhr R. *EMBO J.* 9:3429-3436(1990).

[3] Varghese J.N., Garrett T.P.J., Colman P.M., Chen L., Hoj P.J., Fincher G.B. *Proc. Natl. Acad. Sci. U.S.A.* 91:2785-2789(1994).

244. Glyoxalase I signatures

Glyoxalase I (EC 4.4.1.5) (lactoylglutathione lyase) catalyzes the first step of the glyoxal pathway, the transformation of methylglyoxal and glutathione into S-lactoylglutathione which is then converted by glyoxalase II to lactic acid [1]. Glyoxalase I is an ubiquitous enzyme which binds one mole of zinc per subunit. The bacterial and yeast enzymes are monomeric while the mammalian one is homodimeric. The sequence of glyoxalase I is well conserved. In bacteria and mammals, the enzyme is a protein of about 130 to 180 residues while in fungi it is about twice longer. In these organisms the enzyme is built out of the tandem repeat of an homologous domain. Two signature patterns for this family were derived. The first one is located in the N-terminal region while the second one is located in the central section of the protein and contains a conserved histidine that could be implicated in the binding of the zinc atom.

Consensus pattern: [HQ]-[IVT]-x-[LIVFY]-x-[IV]-x(5)-[STA]-x(2)-F-[YM]-x(2,3)-[LMF]-G-[LMF]-

Consensus pattern: G-[NTKQ]-x(0,5)-[GA]-[LVFY]-[GH]-H-[IVF]-[CGA]-x-[STAGLE]-x(2)-[DNC]-

[1] Kim N.-S., Umezawa Y., Ohmura S., Kato S. J. Biol. Chem. 268:11217-11221(1993).

245. (Glypican)

Glypicans signature

Glypicans [1,2] are a family of heparan sulfate proteoglycans which are anchored to cell membranes by a glycosylphosphatidylinositol (GPI) linkage. Structurally, these proteins consist of three separate domains:

- a) A signal sequence;
- b) An extracellular domain of about 500 residues that contains 12 conserved cysteines probably involved in disulfide bonds and which also contains the

sites of attachment of the heparan sulfate glycosaminoglycan side chains;
 c) A C-terminal hydrophobic region which is post-translationally removed
 after formation of the GPI-anchor.

5 The proteins known to belong to this family are:

- Glypican 1 (GPC1).
- Glypican 2 (GPC2) or cerebroglycan.
- Glypican 3 (GPC3) or OCI-5. In man, defects in GPC3 are the cause of a X-
 10 linked genetic disease, Simpson-Galabi-Behmel syndrome (SGBS).
- K-glypican.
- Glypican 5 (GPC5).
- Drosophila protein dally.

15 The signature pattern that was developed for glypicans is located in the central section of the extracellular domain and contains five of the conserved cysteines.

Consensus pattern C-x(2)-C-x-G-[LIVM]-x(4)-P-C-x(2)-[FY]-C-x(2)-[LIVM]-x(2)-G-C [The
 C's are probably involved in a disulfide bonds] Sequences known to belong to this class
 20 detected by the pattern ALL, except for dally.

[1] Weksberg R., Squire J.A., Templeton D.M. Nat. Genet. 12:225-227(1996).

[2] Watanabe K., Yamada H., Yamaguchi Y. J. Cell Biol. 130:1207-1218(1995).

25

246. Granins signatures

Granins (chromogranins or secretogranins) [1] are a family of acidic proteins present in the
 secretory granules of a wide variety of endocrine and neuro-endocrine cells. The exact
 function(s) of these proteins is not yet known but they seem to be the precursors of
 30 biologically active peptides and/or they may act as helper proteins in the packaging of peptide
 hormones and neuropeptides. Three members of this family of proteins show some sequence
 similarities: - Chromogranin A (CGA) [2]. CGA is a protein of about 420 residues; it is the
 precursor of the peptide pancreastatin which strongly inhibits glucose- induced insulin release

Consensus pattern: [DE]-[SN]-L-[SAN]-x(2)-[DE]-x-E-L-

Consensus pattern: C-[LIVM](2)-E-[LIVM](2)-S-[DN]-[STA]-L-x-K-x-S-x(3)- [LIVM]-[STA]-x-E-C [The two C's are linked by a disulfide bond]-

[11] Huttner W.B., Gerdes H.-H., Rosa P. Trends Biochem. Sci. 16:27-30(1991).

[2] Simon J.-P., Aunis D. Biochem. J. 262:1-13(1989).

247. *grpE* protein signature

In prokaryotes the *grpE* protein [1] stimulates, jointly with *dnaJ*, the ATPase activity of the *dnaK* chaperone. It seems to accelerate the release of ADP from *dnaK* thus allowing *dnaK* to recycle more efficiently. GrpE is a protein of about 22 to 25 Kd. In yeast, an evolutionary related mitochondrial protein (gene GRPE) has been shown [2] to associate with the mitochondrial hsp70 protein and to thus play a role in the import of proteins from the cytoplasm. As a signature pattern, the most conserved region of *grpE* was selected. It is located in the C-terminal section.

Consensus pattern: [FL]-[DN]-[PHEA]-x(2)-[HM]-x-A-[LIVMTN]-x(16,20)-G-[FY]- x(3)-[DEG]-x(2)-[LIVM]-[RI]-x-[SA]-x-V-x-[IV]-

[1] Georgopoulos C., Welch W. *Annu. Rev. Cell Biol.* 9:601-635(1993).

[2] Bolliger L., Deloche O., Glick B.S., Georgopoulos C., Jenoe P., Kronidou N., Horst M., Morishima N., Schatz G. EMBO J. 13:1998-2006(1994).

248. Guanylate kinase signature and profile

Guanylate kinase (EC 2.7.4.8) (GK) [1] catalyzes the ATP-dependent phosphorylation of GMP into GDP. It is essential for recycling GMP and indirectly, cGMP. In prokaryotes (such as *Escherichia coli*), lower eukaryotes (such as yeast) and in vertebrates, GK is a highly conserved monomeric protein of about 200 amino acids. GK has been shown [2,3,4] to be structurally similar to the following proteins: - Protein A57R (or SalG2R) from various strains of *Vaccinia virus*. This protein is highly similar to GK, but contains a frameshift mutation in the N-terminal section and could therefore be inactive in that virus. The following proteins are characterized by the presence in their sequence of one or more copies of the DHR domain, a SH3 domain (see <PDOC50002> as well as a C-terminal GK-like domain, these protein are collectively termed MAGUKs (membrane-associated guanylate kinase homologs) [5]: - *Drosophila lethal(1) discs large-1* tumor suppressor protein (gene *dlg1*). This protein is associated with septate junctions in developing flies and defects in the *dlg1* gene cause neoplastic overgrowth of the imaginal disks. - Mammalian tight junction protein *Zo-1*. - A family of mammalian synaptic proteins that seem to interact with the cytoplasmic tail of NMDA receptor subunits. This family currently consist of SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1 and SAP102. - Vertebrate 55 Kd erythrocyte membrane protein (p55). p55 is a palmitoylated, membrane-associated protein of unknown function. - *Caenorhabditis elegans* protein *lin-2*, which may play a structural role in the induction of the vulva. - Rat protein CASK. - Human protein DLG2. - Human protein DLG3. There is an ATP-binding site (P-loop) in the N-terminal section of GK. This region is not conserved in the GK-like domain of the above proteins which are therefore unlikely to be kinases. However these proteins retain the residues known, in GK, to be involved in the binding of GMP. As a signature pattern a highly conserved region was selected that contains two arginine and a tyrosine which are involved in GMP-binding

Consensus pattern: T-[ST]-R-x(2)-[KR]-x(2)-[DE]-x(2)-G-x(2)-Y-x-[FY]-[LIVMK]-

[1] Stehle T., Schulz G.E. J. Mol. Biol. 224:1127-1141(1992).

[2] Bryant P.J., Woods D.F. Cell 68:621-622(1992).

[3] Goebel M.G. Trends Biochem. Sci. 17:99-99(1992).

[4] Zschocke P.D., Schiltz E., Schulz G.E. Eur. J. Biochem. 213:263-269(1993).

[5] Woods D.F., Bryant P.J. Mech. Dev. 44:85-89(1994).

249. (Glyco_hydro_35)

Glycosyl hydrolases family 35 putative active site

Beta-galactosidases (EC 3.2.1.23) from mammals, fungi, plants and the bacteria *Xanthomonas manihotis* are evolutionary related [1,2]. They belong to family 35 in the classification of glycosyl hydrolases [3,E1].

Mammalian beta-galactosidase is a lysosomal enzyme (gene GLB1) which cleaves the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans and whose deficiency is the cause of the genetic disease Gm(1) gangliosidosis (Morquio disease type B).

On of the best conserved regions in these enzymes contains a glutamic acid residue which, on the basis of similarities with other families of glycosyl hydrolases [4], probably acts as the proton donor in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern: G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY] [The second E is the putative active site residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P. *Glycobiology* 5:603-610(1995).

[2] Carey A.T., Holt K., Picard S., Wilde R., Tucker G.A., Bird C.R., Schuch W., Seymour G.B. *Plant Physiol.* 108:1099-1107(1995).

[3] Henrissat B., Bairoch A. *Biochem. J.* 293:781-788(1993).

[4] Henrissat B., Callebaut I., Fabrega S., Lehn P., Mornon J.-P., Davies G. *Proc. Natl. Acad. Sci. U.S.A.* 92:7090-7094(1995).

250. (Glyco_hydro_16)

Glycosyl hydrolases family 16 signature

It has been shown [1] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- Bacterial beta-1,3-1,4-glucanases, or lichenases, (EC 3.2.1.73) mainly from *Bacillus* but also from *Clostridium thermocellum* (gene *licB*), *Fibrobacter succinogenes* and *Rhodothermus marinus* (gene *bglA*).
- 5 - *Bacillus circulans* beta-1,3-glucanase A1 (EC 3.2.1.39) (gene *glcA*).
- Laminarase (EC 3.2.1.6) from *Clostridium thermocellum* (gene *lam1*).
- *Streptomyces coelicolor* agarase (EC 3.2.1.81) (gene *dagA*).
- *Alteromonas carrageenovora* kappa-carrageenase (EC 3.2.1.83) (gene *cgkA*).

10 Two closely clustered conserved glutamates have been shown [2] to be involved in the catalytic activity of *Bacillus licheniformis* lichenase. The region that contains these residues as a signature pattern was used.

15 Consensus pattern E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA] [The two E's are active site residues]

[1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Juncosa M., Pons J., Dot T., Querol E., Planas A. *J. Biol. Chem.* 269:14530-14535(1994).

20 251. (Glyco_hydro_17)

Glycosyl hydrolases family 17 signature

(aka glycosyl_hydro4)

25 It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- Glucan endo-1,3-beta-glucosidases (EC 3.2.1.39) (endo-(1->3)-beta-glucanase) from
- 30 various plants. This enzyme may be involved in the defense of plants against pathogens through its ability to degrade fungal cell wall polysaccharides.

262

- Glucan 1,3-beta-glucosidase (EC 3.2.1.58) (exo-(1->3)-beta-glucanase) from yeast (gene BGL2). This enzyme may play a role in cell expansion during growth, in cell-cell fusion during mating, and in spore release during sporulation.

- Lichenases (EC 3.2.1.73) (endo-(1->3,1->4)-beta-glucanase) from various plants.

5

The best conserved region in the sequence of these enzymes is located in their central section. This region contains a conserved tryptophan residue which could be involved in the interaction with the glucan substrates [2] and it also contains a conserved glutamate which has been shown [3] to act as the nucleophile in the catalytic mechanism. This region was used as a signature pattern.

10

Consensus pattern [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ]
[E is an active site residue] Sequences known to belong to this class detected by the pattern ALL.

15

[1] Henrissat B. Biochem. J. 280:309-316(1991).

[2] Ori N., Sessa G., Lotan T., Himmelhoch S., Fluhr R. EMBO J. 9:3429-3436(1990).

[3] Varghese J.N., Garrett T.P.J., Colman P.M., Chen L., Hoj P.J., Fincher G.B. Proc. Natl. Acad. Sci. U.S.A. 91:2785-2789(1994).

20

252. (Glyco_hydro_3)

Glycosyl hydrolases family 3 active site

25

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Beta glucosidases (EC 3.2.1.21) from the fungi *Aspergillus wentii* (A-3), *Hansenula anomala*, *Kluyveromyces fragilis*, *Saccharomycopsis fibuligera*, (BGL1 and BGL2), *Schizophyllum commune* and *Trichoderma reesei* (BGL1).
- Beta glucosidases from the bacteria *Agrobacterium tumefaciens* (Cbg1), *Butyrivibrio fibrisolvens* (bglA), *Clostridium thermocellum* (bglB), *Escherichia coli* (bglX), *Erwinia chrysanthemi* (bgxA) and *Ruminococcus*

30

albus.

- Alteromonas strain O-7 beta-hexosaminidase A (EC 3.2.1.52).
 - Bacillus subtilis hypothetical protein yzbA.
 - Escherichia coli hypothetical protein ycfO and HI0959, the corresponding
- 5 Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in Aspergillus wentii beta- glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

10 Consensus pattern[LIVM](2)-[KR]-x-[EOK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]- [ST]-D-x(2)-[SGADNI] [D is the active site residue] Sequences known to belong to this class detected by the patternALL.

- 15 [1] Henrissat B. Biochem. J. 280:309-316(1991).
- [2] Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).
- [3] Bause E., Legler G. Biochim. Biophys. Acta 626:459-465(1980).

20 253. (Glyco_hydro_28)

Polygalacturonase active site (aka PG)

Polygalacturonase (EC 3.2.1.15) (PG) (pectinase) [1,2] catalyzes the random hydrolysis of 1,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans. In fruit,

25 polygalacturonase plays an important role in cell wall metabolism during ripening. In plant bacterial pathogens such as Erwinia carotovora or Pseudomonas solanacearum and fungal pathogens such as Aspergillus niger, polygalacturonase is involved in maceration and soft-rotting of plant tissue.

30 Exo-poly-alpha-D-galacturonosidase (EC 3.2.1.82) (exoPG) [3] hydrolyzes peptic acid from the non-reducing end, releasing digalacturonate.

Prokaryotic, eukaryotic PG and exoPG share a few regions of sequence similarity. The best conserved of these regions was selected. It is centered on a conserved histidine most probably involved in the catalytic mechanism [4].

- 5 Consensus pattern[GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)-[LIVM]-G-S [H is the putative active site residue] Sequences known to belong to this class detected by the patternALL.

Note: these proteins belong to family 28 in the classification of glycosyl hydrolases [5].

10

[1] Ruttowski E., Labitzke R., Khanh N.Q., Loeffler F., Gottschalk M., Jany K.-D. Biochim. Biophys. Acta 1087:104-106(1990).

[2] Huang J., Schell M.A. J. Bacteriol. 172:3879-3887(1990).

[3] He S.Y., Collmer A. J. Bacteriol. 172:4988-4995(1990).

15

[4] Bussink H.J.D., Buxton F.P., Visser J. Curr. Genet. 19:467-474(1991).

[5] Henrissat B. Biochem. J. 280:309-316(1991).

254. (Glyco_hydro_32)

20

Glycosyl hydrolases family 32 active site

It has been shown [1,2] that the following glycosyl hydrolases can be classified into a single family on the basis of sequence similarities:

- 25 - Inulinase (EC 3.2.1.7) (or inulase) from the fungi *Kluyveromyces marxianus*.
 - Beta-fructofuranosidase (EC 3.2.1.26), commonly known as invertase in fungi and plants and as sucrase in bacteria (gene *sacA* or *scrB*).
 - Raffinose invertase (EC 3.2.1.26) (gene *rafD*) from *Escherichia coli* plasmid pRSD2.
 30 - Levanase (EC 3.2.1.65) (gene *sacC*) from *Bacillus subtilis*.

One of the conserved regions in these enzymes is located in the N-terminal section and contains an aspartic acid residue which has been shown [3], in yeast invertase to be important for the catalytic mechanism. This region was used as a signature pattern.

- 5 Consensus pattern H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G [D is the active site residue]
Sequences known to belong to this class detected by the patternALL.

[1] Henrissat B. Biochem. J. 280:309-316(1991).

[2] Gunasekaran P., Karunakaran T., Cami B., Mukundan A.G., Preziosi L., Baratti J. J.

10 Bacteriol. 172:6727-6735(1990).

[3] Reddy V.A., Maley F. J. Biol. Chem. 265:10817-10120(1990).

255. (Glyco_hydro_1)

15 Glycosyl hydrolases family 1 signatures

It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- 20 - Beta-glucosidases (EC 3.2.1.21) from various bacteria such as *Agrobacterium* strain ATCC 21400, *Bacillus polymyxa*, and *Caldocellum saccharolyticum*.
- Two plants (clover) beta-glucosidases (EC 3.2.1.21).
- Two different beta-galactosidases (EC 3.2.1.23) from the archaeobacteria *Sulfolobus solfataricus* (genes *bgaS* and *lacS*).
25 - 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as *Lactobacillus casei*, *Lactococcus lactis*, and *Staphylococcus aureus*.
- 6-phospho-beta-glucosidases (EC 3.2.1.86) from *Escherichia coli* (genes *bglB* and *ascB*) and from *Erwinia chrysanthemi* (gene *arbB*).
- Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase).
30 - Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62).
LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is

evolutionary related to the above glycosyl hydrolases.

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-glucosidase from *Agrobacterium*, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. This region was used as a signature pattern. As a second signature pattern we selected a conserved region, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue.

Consensus pattern[LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue] Sequences known to belong to this class detected by the patternALL.

Note: this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].

Consensus patternF-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA] Sequences known to belong to this class detected by the pattern ALL.

Note: this pattern will pick up the last three domains of LPH.

[1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Henrissat B. *Protein Seq. Data Anal.* 4:61-62(1991).

[3] Gonzalez-Candelas L., Ramon D., Polaina J. *Gene* 95:31-38(1990).

[4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. *Bacteriol.* 174:765-777(1992).

[5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. *Am. Chem. Soc.* 112:5887-5889(1990).

256. Glyco_hydro_20

Glycosyl hydrolase family 20

Previous Pfam IDs: glycosyl_hydr11;

Number of members: 33

257. (Glyco_hydro_9)

Glycosyl hydrolases family 9 active sites signatures

5 (aka Glycosyl_hydr12)

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family E [3] or as the glycosyl hydrolases family 9 [4,E1]. The enzymes which are currently known to belong to this family are listed below.

- 15 - *Butyrivibrio fibrisolvens* cellodextrinase I (ced1).
- *Cellulomonas fimi* endoglucanases B (cenB) and C (cenC).
- *Clostridium cellulolyticum* endoglucanase G (celCCG).
- *Clostridium cellulovorans* endoglucanase C (engC).
- *Clostridium stercoararium* endoglucanase Z (avicelase I) (celZ).
- 20 - *Clostridium thermocellum* endoglucanases D (celD), F (celF) and I (celI).
- *Fibrobacter succinogenes* endoglucanase A (endA).
- *Pseudomonas fluorescens* endoglucanase A (celA).
- *Streptomyces reticuli* endoglucanase I (cel1).
- *Thermomonospora fusca* endoglucanase E-4 (celD).
- 25 - *Dictyostelium discoideum* spore germination specific endoglucanase 270-6. This slime mold enzyme may digest the spore cell wall during germination, to release the enclosed amoeba.
- Endoglucanases from plants such as Avocado or French bean. In plants this enzyme may be
- 30 involved the fruit ripening process.

Two of the most conserved regions in these enzymes are centered on conserved residues which have been shown [5,6], in the endoglucanase D from *Cellulomonas thermocellum*, to

be important for the catalytic activity. The first region contains an active site histidine and the second region contains two catalytically important residues: an aspartate and a glutamate. Both regions were used as signature patterns.

- 5 Consensus pattern [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R [H is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for *Cellulomonas fimi* cenC and *Streptomyces reticuli* cell1.

- 10 Consensus pattern [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA] [D and E are active site residues] Sequences known to belong to this class detected by the pattern ALL, except for *Fibrobacter succinogenes* endA whose sequence seems to be incorrect.

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

- 15 [2] Gilkes N.R., Henriissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henriissat B., Claeysens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henriissat B. Biochem. J. 280:309-316(1991).

[5] Tomme P., Chauvaux S., Beguin P., Millet J., Aubert J.-P., Claeysens M. J. Biol. Chem. 266:10313-10318(1991).

- 20 [6] Tomme P., van Beeumen J., Claeysens M. Biochem. J. 285:319-324(1992).

258. Matrix protein (MA), p15 (GAG_ma)

The matrix protein, p15, is encoded by the gag gene. MA is involved in pathogenicity

- 25 [1].

[1] : Pozsgay JM, Beilharz MW, Wines BD, Hess AD, Pitha PM, J Virol 1993;67:5989-5999.

- 30 259. Gag polyprotein, inner coat protein p12 (GAG_P12)

The retroviral p12 is a virion structural protein. p12 is proline rich. The function carried out by p12 in assembly and replication is unknown. p12C is associated with pathogenicity of the virus

[1] Pozsgay JM, Beilharz MW, Wines BD, Hess AD, Pitha PM, J Virol 1993;67:5989-5999.

260. Glutamine synthetase signatures (GLN-SYNT)

- 5 Glutamine synthetase (EC 6.3.1.2) (GS) [1] plays an essential role in the metabolism of nitrogen by catalyzing the condensation of glutamate and ammonia to form glutamine. There seem to be three different classes of GS [2,3,4]: - Class I enzymes (GSI) are specific to prokaryotes, and are oligomers of 12 identical subunits. The activity of GSI-type enzyme is controlled by the adenylation of a tyrosine residue. The adenylylated enzyme is inactive. -
- 10 Class II enzymes (GSII) are found in eukaryotes and in bacteria belonging to the Rhizobiaceae, Frankiaceae, and Streptomycetaceae families (these bacteria have also a class-I GS). GSII are octamer of identical subunits. Plants have two or more isozymes of GSII, one of the isozymes is translocated into the chloroplast. - Class III enzymes (GSIII) has, currently, only been found in *Bacteroides fragilis* and in *butyrivibrio fibrisolvens*. It is a hexamer of identical chains. It is much larger (about 700 amino acids) than the GSI (450 to
- 15 470 amino acids) or GSII (350 to 420 amino acids) enzymes. While the three classes of GS's are clearly structurally related, the sequence similarities are not so extensive. As signature patterns three conserved regions were selected. The first pattern is based on a conserved tetrapeptide in the N-terminal section of the enzyme, the second one is based on a glycine-rich region which is thought to be involved in ATP-binding. The third pattern is specific to
- 20 class I glutamine synthetases and includes the tyrosine residue which is reversibly adenylylated.

Consensus pattern: [FYWL]-D-G-S-S-x(6,8)-[DENOSTAK]-[SA]-[DE]-x(2)-[LIVMFY]-

- 25 Consensus pattern: K-P-[LIVMFYA]-x(3,5)-[NPAT]-G-[GSTAN]-G-x-H-x(3)-S-

Consensus pattern: K-[LIVM]-x(5)-[LIVMA]-D-[RK]-[DN]-[LI]-Y [Y is the site of adenylation]-

- [1] Eisenberg D., Almasy R.J., Janson C.A., Chapman M.S., Suh S.W., Cascio D., Smith W.W. Cold Spring Harbor Symp. Quant. Biol. 52:483-490(1987).

[2] Kumada Y., Benson D.R., Hillemann D., Hosted T.J., Rochefort D.A., Thompson C.J., Wohlleben W., Tateno Y. Proc. Natl. Acad. Sci. U.S.A. 90:3009-3013(1993).

[3] Shatters R.G., Kahn M.L. J. Mol. Evol. 29:422-428(1989).

[4] Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F. J. Mol. Evol. 38:566-576(1994).

261. Globins profile (globin1)

- 5 Globins are heme-containing proteins involved in binding and/or transporting oxygen [1]. They belong to a very large and well studied family which is widely distributed in many organisms. The major groups of globins are: - Hemoglobins (Hb) from vertebrates. Hb is the protein responsible for transporting oxygen from the lungs to other tissues. It is a tetramer of two alpha and two beta chains. Most vertebrate species also express specific embryonic or fetal forms of hemoglobin where the alpha or the beta chains are replaced by a chain with higher oxygen affinity, as for the gamma, delta, epsilon and zeta chains in mammals, for example. - Myoglobins (Mg) from vertebrates. Mg is a monomeric protein responsible for oxygen storage in muscles. - Invertebrate globins [2]. A wide variety of globins are found in invertebrates. Molluscs generally have one or two muscle globins which are either monomeric or dimeric. Insects, such as the midge *Chironomus thummi*, have a large set of extracellular globins. Nematodes and annelids have a variety of intracellular and extracellular globins; some of them are multi- domain polypeptides (from two up to nine-domain globins) and some produce large, disulfide-bonded aggregates. - Leghemoglobins (Lg) from the root nodules of leguminous plants. Lg provides oxygen for bacteroids. - Flavohemoproteins from bacteria (*Escherichia coli hmpA*) and fungi [3]. These proteins consist of two distinct domains: an N-terminal globin domain and a C-terminal FAD-containing reductase domain. In bacteria such as *Vitreoscilla*, the enzyme-associated globin is a single domain protein. All these globins seem to have evolved from a common ancestor. The profile developed to detect members of the globin family is based on a structural alignment of selected globin sequences
- 25 [1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).[2] Goodman M., Pedwaydon J., Czelusniak J., Suzuki T., Gotoh T., Moens L., Shishikura F., Walz D., Vinogradov S. J. Mol. Evol. 27:236-249(1988).

Plant hemoglobins signature (globin2)

- 30 Leghemoglobins [1] are hemoproteins present in the root nodules of leguminous plants. Leghemoglobins are structurally and functionally related to hemoglobin and myoglobin. By providing oxygen to the bacteroids, they are essential for symbiotic nitrogen fixation. Structurally related hemoglobins from the nodules of non-leguminous plants [2,3], and from

the roots of non-nodulating plants[4] have been recently sequenced. A signature pattern was developed that picks up the sequence of plants hemoglobins, exclusively.

Consensus pattern: [SN]-P-x-L-x(2)-H-A-x(3)-F-

[1] Powell R., Gannon F. BioEssays 9:117-121(1988).

[2] Kortt A.A., Trinick M.J., Appleby C.A. Eur. J. Biochem. 175:141-149(1988).

[3] Kortt A.A., Inglis A.S., Fleming A.I., Appleby C.A. FEBS Lett. 231:341-346(1988).

[4] Bogusz D., Appleby C.A., Landsmann J., Dennis E.S., Trinick M.J., Peacock W.J.

Nature 331:178-180(1988).

262. Fructose-bisphosphate aldolase class-I active site (glycolytic_enz)

Fructose-bisphosphate aldolase [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6-bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate. There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-I aldolases [3], mainly found in higher eukaryotes, are homotetrameric enzymes which form a Schiff-base intermediate between the C-2 carbonyl group of the substrate (dihydroxyacetone phosphate) and the epsilon-amino group of a lysine residue. In vertebrates, three forms of this enzyme are found: aldolase A in muscle, aldolase B in liver and aldolase C in brain. The sequence around the lysine involved in the Schiff-base is highly conserved and can be used as a signature for this class of enzyme.

Consensus pattern: [LIVM]-x-[LIVMFYW]-E-G-x-[LS]-L-K-P-[SN] [K is involved in Schiff-base formation]-

[1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).

[2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).

[3] Freemont P.S., Dunbar B., Fothergill-Gilmore L.A. Biochem. J. 249:779-788(1988).

263. Glycosyl hydrolases family 11 active sites signatures

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family G [3] or as the glycosyl hydrolases family 11 [4,E1]. The enzymes which are currently known to belong to this family are listed below. - *Aspergillus awamori* xylanase C (xynC). - *Bacillus circulans*, *pumilus*, *stearothermophilus* and *subtilis* xylanase (xynA). - *Clostridium acetobutylicum* xylanase (xynB). - *Clostridium stercorarium* xylanase A (xynA). - *Fibrobacter succinogenes* xylanase C (xynC) which consist of two catalytic domains that both belong to family 10. - *Neocallimastix patriciarum* xylanase A (xynA). - *Ruminococcus flavefaciens* bifunctional xylanase XYLA (xynA). This protein consists of three domains: a N-terminal xylanase catalytic domain that belongs to family 11 of glycosyl hydrolases; a central domain composed of short repeats of Gln, Asn an Trp, and a C-terminal xylanase catalytic domain that belongs to family 10 of glycosyl hydrolases. - *Schizophyllum commune* xylanase A. - *Streptomyces lividans* xylanases B (xlnB) and C (xlnC). - *Trichoderma reesei* xylanases I and II. Two of the conserved regions in these enzymes are centered on glutamic acidresidues which have both been shown [5], in *Bacillus pumilis* xylanase, to be necessary for catalytic activity. Both regions were used as signature patterns.

Consensus pattern: [PSA]-[LQ]-x-E-Y-Y-[LIVM](2)-[DE]-x-[FYWHN] [E is an active site residue]-

Consensus pattern: [LIVMF]-x(2)-E-[AG]-[YWG]-[QRFGS]-[SG]-[STAN]-G-x-[SAF] [E is an active site residue]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henriissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henriissat B., Claeysens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henriissat B. Biochem. J. 280:309-316(1991).

[5] Ko E.P., Akatsuka H., Moriyama H., Shinmyo A., Hata Y., Katsube Y., Urabe I., Okada H. Biochem. J. 288:117-121(1992).

264. Glycosyl hydrolase family 14

This family are beta amylases.

265. Glycosyl hydrolases family 1 signatures

It has been shown [1 to 4] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta-glucosidases (EC 3.2.1.21) from various bacteria such as *Agrobacterium* strain ATCC 21400, *Bacillus polymyxa*, and *Caldocellum saccharolyticum*. - Two plants (clover) beta-glucosidases (EC 3.2.1.21). - Two different beta-galactosidases (EC 3.2.1.23) from the archaeobacteria *Sulfolobus solfataricus* (genes *bgaS* and *lacS*). - 6-phospho-beta-galactosidases (EC 3.2.1.85) from various bacteria such as *Lactobacillus casei*, *Lactococcus lactis*, and *Staphylococcus aureus*. - 6-phospho-beta-glucosidases (EC 3.2.1.86) from *Escherichia coli* (genes *bglB* and *ascB*) and from *Erwinia chrysanthemi* (gene *arbB*). - Plants myrosinases (EC 3.2.3.1) (sinigrinase) (thioglucosidase). - Mammalian lactase-phlorizin hydrolase (LPH) (EC 3.2.1.108 / EC 3.2.1.62). LPH, an integral membrane glycoprotein, is the enzyme that splits lactose in the small intestine. LPH is a large protein of about 1900 residues which contains four tandem repeats of a domain of about 450 residues which is evolutionary related to the above glycosyl hydrolases. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the beta-glucosidase from *Agrobacterium*, to be directly involved in glycosidic bond cleavage by acting as a nucleophile. This region was used as a signature pattern. As a second signature pattern a conserved region was selected, found in the N-terminal extremity of these enzymes, this region also contains a glutamic acid residue.

Consensus pattern: [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] [E is the active site residue]

Note: this pattern will pick up the last two domains of LPH; the first two domains, which are removed from the LPH precursor by proteolytic processing, have lost the active site glutamate and may therefore be inactive [4].

Consensus pattern: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA]-

- [1] Henrissat B. Biochem. J. 280:309-316(1991).
 [2] Henrissat B. Protein Seq. Data Anal. 4:61-62(1991).
 [3] Gonzalez-Candelas L., Ramon D., Polaina J. Gene 95:31-38(1990).
 5 [4] El Hassouni M., Henrissat B., Chippaux M., Barras F. J. Bacteriol. 174:765-777(1992).
 [5] Withers S.G., Warren R.A.J., Street I.P., Rupitz K., Kempton J.B., Aebersold R. J. Am. Chem. Soc. 112:5887-5889(1990).

10 266. Glycosyl hydrolases family 2 signatures

It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family: - Beta-galactosidases (EC 3.2.1.23) from bacteria such as Escherichia coli (genes lacZ and ebgA), Clostridium acetobutylicum, Clostridium thermosulfurogenes, Klebsiella pneumoniae, Lactobacillus delbrueckii, or
 15 Streptococcus thermophilus and from the fungi Kluyveromyces lactis. - Beta-glucuronidase (EC 3.2.1.31) from Escherichia coli (gene uidA) and from mammals. One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in Escherichia coli lacZ, to be the general acid/base catalyst in the active site of the enzyme. This region was used as a signature pattern. As a second signature pattern a
 20 highly conserved region was selected located some sixty residues upstream from the active site glutamate.

Consensus pattern: N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4)-

- 25 Consensus pattern: [DENQLF]-[KRVW]-N-[HRY]-[STAPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue]-

- [1] Henrissat B. Biochem. J. 280:309-316(1991).
 [2] Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. Gen. Microbiol.
 30 137:369-380(1991).
 [3] Gebler J.C., Aebersold R., Withers S.G. J. Biol. Chem. 267:11126-11130(1992).

267. Glycosyl hydrolases family 3 active site

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- Beta glucosidases (EC 3.2.1.21) from the fungi *Aspergillus wentii* (A-3),
 5 *Hansenula anomala*, *Kluyveromyces fragilis*, *Saccharomycopsis fibuligera*,
 (BGL1 and BGL2), *Schizophyllum commune* and *Trichoderma reesei* (BGL1).
- Beta glucosidases from the bacteria *Agrobacterium tumefaciens* (Cbgl),
Butyrivibrio fibrisolvens (bglA), *Clostridium thermocellum* (bglB),
Escherichia coli (bglX), *Erwinia chrysanthemi* (bgxA) and *Ruminococcus*
 10 *albus*. - *Alteromonas* strain O-7 beta-hexosaminidase A (EC 3.2.1.52).
- *Bacillus subtilis* hypothetical protein yzbA.
- *Escherichia coli* hypothetical protein ycfO and HI0959, the corresponding
Haemophilus influenzae protein.

One of the conserved regions in these enzymes is centered on a conserved
 15 aspartic acid residue which has been shown [3], in *Aspergillus wentii* beta-
 glucosidase A3, to be implicated in the catalytic mechanism. This
 region was used as a signature pattern.

Consensus pattern: [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-
 20 x(2)-[SGADNI] [D is the active site residue]

[1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Castle L.A., Smith K.D., Morris R.O. *J. Bacteriol.* 174:1478-1486(1992).

[3] Bause E., Legler G. *Biochim. Biophys. Acta* 626:459-465(1980).

268. Glycosyl hydrolases family 8 signature

The microbial degradation of cellulose and xylans requires several types of enzymes such as
 endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91)(exoglucanases), or xylanases
 30 (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes
 (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into
 families. One of these families is known as the cellulase family D [3] or as the glycosyl
 hydrolases family 8 [4,E1]. The enzymes which are currently known to belong to this family

are listed below. - *Acetobacter xylinum* endonuclease cmcAX. - *Bacillus* strain KSM-330 acidic endonuclease K (Endo-K). - *Cellulomonas josui* endoglucanase 2 (celB). - *Cellulomonas uda* endoglucanase. - *Clostridium cellulolyticum* endoglucanases C (celcCC). - *Clostridium thermocellum* endoglucanases A (celA). - *Erwinia chrysanthemi* minor endoglucanase y (celY). - *Bacillus circulans* beta-glucanase (EC 3.2.1.73). - *Escherichia coli* hypothetical protein yhjM. The most conserved region in these enzymes is a stretch of about 20 residues that contains two conserved aspartate. The first aspartate is thought [5] to act as the nucleophile in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern: A-[ST]-D-[AG]-D-x(2)-[IM]-A-x-[SA]-[LIVM]-[LIVMG]-x-A- x(3)-[FW] [The first D is an active site residue]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeyssens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

[5] Alzari P.M., Souchon H., Dominguez R. Structure 4:265-275(1996).

269. Glycosyl hydrolases family 9 active sites signatures

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produce a spectrum of cellulolytic enzymes (cellulases)

and xylanases which, on the basis of sequence similarities, can be classified into families.

One of these families is known as the cellulase family E [3] or as the glycosyl hydrolases family 9 [4,E1]. The enzymes which are currently known to belong to this family are listed below.

- *Butyrivibrio fibrisolvens* cellobextrinase 1 (ced1). - *Cellulomonas fimi* endoglucanases B (cenB) and C (cenC). - *Clostridium cellulolyticum* endoglucanase G (celCCG). - *Clostridium cellulovorans* endoglucanase C (engC). - *Clostridium stercoarium* endoglucanase Z (avicelase I) (celZ). - *Clostridium thermocellum* endoglucanases D (celD), F (celF) and I (celI). - *Fibrobacter succinogenes* endoglucanase A (endA). - *Pseudomonas fluorescens* endoglucanase A (celA). - *Streptomyces reticuli* endoglucanase 1 (cel1). -

277

Thermomonospora fusca endoglucanase E-4 (celD). - *Dictyostelium discoideum* spore germination specific endoglucanase 270-6. This slime mold enzyme may digest the spore cell wall during germination, to release the enclosed amoeba. - Endoglucanases from plants such as Avocado or French bean. In plants this enzyme may be involved the fruit ripening process.

Two of the most conserved regions in these enzymes are centered on conserved residues which have been shown [5,6], in the endoglucanase D from *Cellulomonas thermocellum*, to be important for the catalytic activity. The first region contains an active site histidine and the second region contains two catalytically important residues: an aspartate and a glutamate. Both regions were used as signature patterns.

Consensus pattern: [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R [H is an active site residue]-

Consensus pattern: [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA] [D and E are active site residues]-

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

[2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeysens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

[5] Tomme P., Chauvaux S., Beguin P., Millet J., Aubert J.-P., Claeysens M. J. Biol. Chem. 266:10313-10318(1991).

[6] Tomme P., van Beeumen J., Claeysens M. Biochem. J. 285:319-324(1992).

270. Glyceraldehyde 3-phosphate dehydrogenase active site (gpdh)

Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH) [1] is a tetrameric NAD-binding enzyme common to both the glycolytic and gluconeogenic pathways. A cysteine in the middle of the molecule is involved in forming a covalent phosphoglycerol thioester intermediate. The sequence around this cysteine is totally conserved in eubacterial and eukaryotic GAPDHs and is also present, albeit in a variant form, in the otherwise highly divergent archaeobacterial GAPDH [2]. *Escherichia coli* D-erythrose 4-phosphate dehydrogenase (E4PDH) (gene *epd* *orgapB*) is an enzyme highly related to GAPDH [3].

Consensus pattern: [ASV]-S-C-[NT]-T-x(2)-[LIM] [C is the active site residue]-

[1] Harris J.L., Waters M. (In) The Enzymes (3rd edition) 13:1-50(1976).

5 [2] Fabry S., Lang J., Niermann T., Vingron M., Hensel R. Eur. J. Biochem. 179:405-413(1989).

[3] Zhao G., Pease A.J., Bharani N., Winkler M.E. J. Bacteriol. 177:2804-2812(1995).

10 271. Granulins signature

Granulins [1] are a family of cysteine-rich peptides of about 6 Kd which may have multiple biological activity. A precursor protein (known as acrogranin) potentially encodes seven different forms of granulin (grnA to grnG) which are probably released by post-translational proteolytic processing. A schematic representation of the structure of a granulin is shown below: xxxCxxxxxCxxxxCCxxxxxxxxCCxxxxxCxxxxxCxxxxxCxxxxxCx
 15 *****C*: conserved cysteine probably involved in a disulfide bond.*': position of the pattern. Granulins are evolutionary related to a PMP-D1, a peptide extracted from the pars intercerebralis of migratory locusts [2].

20 Consensus pattern: C-x-D-x(2)-H-C-C-P-x(4)-C [The four C's are probably involved in disulfide bonds]-

[1] Bhandari V., Palfree R.G., Bateman A. Proc. Natl. Acad. Sci. U.S.A. 89:1715-1719(1992).

25 [2] Nakakura N., Hietter H., van Dorselaer A., Luu B. Eur. J. Biochem. 204:147-153(1992).

272. (HCV RdRp) Hepatitis C virus RNA dependent RNA polymerase

30 The RNA dependent RNA polymerase is also known as non-structural protein NS5B. NS5B is a 65 kDa protein that resembles other viral RNA polymerases. HCV replication is thought to occur in membrane bound replication

complexes. These complexes transcribe the positive strand and the resulting minus strand is used as a template for the synthesis of genomic RNA. There are two viral proteins involved in the reaction, NS3 and NS5B.[1,2]

- [1] Lohmann V, Korner F, Herian U, Bartenschlager R; J Virol 1997;71:8416-8428. [2] Behrens SE, Tomei L, De Francesco R; EMBO J 1996;15:12-22. [3] Ishido S, Fujita T, Hotta H; Biochem Biophys Res Commun 1998;244:35-40.

273. (HHH) Helix-hairpin-helix motif.

- [1] Doherty AJ, Serpell LC, Ponting CP; Nucleic Acids Res 1996;24:2488-2497.

274. HIT family signature

Recently a family of small proteins of about 12 to 16 Kd has been described[1]. This family currently consists of: - Mammalian protein HINT (also known as Protein kinase C inhibitor 1 or PKCI- 1). HINT was incorrectly thought to be a specific inhibitor of PKC. It has been shown to bind zinc. - Fission yeast diadenosine 5',5'''-P1,P4-tetraphosphate asymmetrical hydrolase (Ap4Aase) (EC 3.6.1.17) [2] (gene aph1), which cleaves A-5'-PPPP- 5'A to yield AMP and ATP. - FHIT, a human protein whose gene is altered in different tumors and which acts [3] as a diadenosine 5',5'''-P1,P3-triphosphate hydrolase (Ap3Aase) (EC 3.6.1.29) cleaving A-5'-PPP-5'A to yield AMP and ADP. - Yeast proteins HNT1 and HNT2. - Maize zinc-binding protein ZBP14. - Escherichia coli hypothetical protein ycfF. - Haemophilus influenzae hypothetical protein HI0961. - Helicobacter pylori hypothetical protein HP0404. - Methanococcus jannaschii hypothetical protein MJ0866. - Mycobacterium leprae hypothetical protein U296A. - Synechocystis strain PCC 6803 hypothetical protein slr1234. - Caenorhabditis elegans hypothetical protein F21C3.3. - A hypothetical 13.2 Kd protein in hisE 3'region in Azospirillum brasilense. - A hypothetical 13.1 Kd protein in p37 5'region in Mycoplasma hyorhinitis. - A hypothetical 12.4 Kd protein in psbAII 5'region in Synechococcus strain PCC 7942. All these proteins contains a region with three clustered

histidines. This region is responsible for the designation of this family: HIT, for 'HistidineTriad' [1]. This region was originally thought to be implied in the binding of a zinc ion but was later identified [4] as part of the alpha-phosphate binding site of a nucleotide-binding domain. As a signature pattern, the region of the histidine triad was selected.

Consensus pattern: [NQA]-x(4)-[GAV]-x-[QF]-x-[LIVM]-x-H-[LIVMFYT]-H-[LIVMFT]-H-[LIVMF](2)-[PSGA]-

[1] Seraphin B. DNA Seq. 3:177-179(1992).

[2] Huang Y., Garrison P.N., Barnes L.D. Biochem. J. 312:925-932(1995).

[3] Barnes L.D., Garrison P.N., Siprashvili Z., Guranowski A., Robinson A.K., Ingram S.W., Croce C.M., Ohta M., Huebner K. Biochemistry 35:11529-11535(1996).

[4] Brenner C., Garrison P., Gilmour J., Peisach D., Ringe D., Petsko G.A., Lowenstein J.M. Nat. Struct. Biol. 4:231-238(1997).

275. Myc-type, 'helix-loop-helix' dimerization domain signature (HLH)

A number of eukaryotic proteins, which probably are sequence specific DNA-binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are referred as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also referred to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW. - The myc

family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation. - Proteins involved in myogenesis (the induction of muscle cells). In mammals MyoD1 (Myf-3), myogenin (Myf-4), Myf-5, and Myf-6 (Mrf4 or herculin), in birds CMD1 (OMF-1), in *Xenopus* MyoD and MF25, in *Caenorhabditis elegans* CeMyoD, and in *Drosophila nautilus* (nau). - Vertebrate proteins that bind specific DNA sequences ('E boxes') in various immunoglobulin chains enhancers: E2A or ITF-1 (E12/pan-2 and E47/pan-1), ITF-2 (tcf4), TFE3, and TFEB. - Vertebrate neurogenic differentiation factor 1 that acts as differentiation factor during neurogenesis. - Vertebrate MAX protein, a transcription regulator that forms a sequence- specific DNA-binding protein complex with myc or mad. - Vertebrate Max Interacting Protein 1 (MXI1 protein) which acts as a transcriptional repressor and may antagonize myc transcriptional activity by competing for max. - Proteins of the bHLH/PAS superfamily which are transcriptional activators. In mammals, AH receptor nuclear translocator (ARNT), single-minded homologs (SIM1 and SIM2), hypoxia-inducible factor 1 alpha (HIF1A), AH receptor (AHR), neuronal pas domain proteins (NPAS1 and NPAS2), endothelial pas domain protein 1 (EPAS1), mouse ARNT2, and human BMAL1. In *Drosophila*, single-minded (SIM), AH receptor nuclear translocator (ARNT), tracheless protein (TRH), and similar protein (SIMA). - Mammalian transcription factors HES, which repress transcription by acting on two types of DNA sequences, the E box and the N box. - Mammalian MAD protein (max dimerizer) which acts as transcriptional repressor and may antagonize myc transcriptional activity by competing for max. - Mammalian Upstream Stimulatory Factor 1 and 2 (USF1 and USF2), which bind to a symmetrical DNA sequence that is found in a variety of viral and cellular promoters. - Human lyl-1 protein; which is involved, by chromosomal translocation, in T- cell leukemia. - Human transcription factor AP-4. - Mouse helix-loop-helix proteins MATH-1 and MATH-2 which activate E box- dependent transcription in collaboration with E47. - Mammalian stem cell protein (SCL) (also known as tal1), a protein which may play an important role in hemopoietic differentiation. SCL is involved, by chromosomal translocation, in stem-cell leukemia. - Mammalian proteins Id1 to Id4 [5]. Id (inhibitor of DNA binding) proteins lack a basic DNA-binding domain but are able to form heterodimers with other HLH proteins, thereby inhibiting binding to DNA. - *Drosophila* extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins. - Human Sterol

Drosophila atonal protein (ato) which is involved in neurogenesis. - Drosophila daughterless (da) protein, which is essential for neurogenesis and sex-determination. - Drosophila deadpan (dpn), a hairy-like protein involved in the functional differentiation of neurons. - Drosophila

E(spl), that are hairy-like proteins active during neurogenesis, also act as transcriptional repressors. - *Drosophila twist (twi)* protein, which is involved in the establishment of germ

Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression. -

Amphipathic helix 1 Loop Amphipathic helix 2. The signature pattern developed to detect this domain spans completely the second amphipathic helix.

Consensus pattern: [DENSTAP]-[KTR]-[LIVMAGSNT]-{FYWCPHKR}-[LIVMT]-
[LIVM]-x(2)-[STAV]-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]-{P}-{P}-
[LIVMRKHQ]-

- [1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).
[2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).

[3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).

[4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990).

[5] Riechmann V., van Cruechten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

276. HMG14 and HMG17 signature

High mobility group (HMG) proteins are a family of relatively low molecular weight non-histone components in chromatin. HMG14 and HMG17 [1], two related proteins of about 100 amino acid residues, bind to the inner side of the nucleosomal DNA thus altering the interaction between the DNA and the histone octamer. These two proteins may be involved in the process which maintains transcribable genes in a unique chromatin conformation. The trout nonhistone chromosomal protein H6 (histone T) also belongs to this family. As a signature pattern a conserved stretch of 10 residues located in the N-terminal section of HMG14 and HMG17 was selected.

Consensus pattern: R-R-S-A-R-L-S-A-[RK]-P-

[1] Bustin M., Reeves R. Prog. Nucleic Acid Res. Mol. Biol. 54:35-100(1996).

277. Hydroxymethylglutaryl-coenzyme A lyase active site (HMGL1)

3-hydroxy-3-methylglutaryl-coenzyme A lyase (HMG-CoA lyase or HL) (EC 4.1.3.4) catalyzes the transformation of HMG-CoA into acetyl-CoA and acetoacetate. In vertebrates it is a mitochondrial enzyme which is involved in ketogenesis and in leucine catabolism [1]. In some bacteria, such as *Pseudomonas mevalonii*, it is involved in mevalonate catabolism (gene *mvaB*). A cysteine has been shown [2], in *mvaB*, to be required for the activity of the enzyme. The region around this residue is perfectly conserved and is used as a signature pattern.

Consensus pattern: S-V-A-G-L-G-G-C-P-Y [C is the active site residue]-

[1] Mitchell G.A., Robert M.-F., Hruz P.W., Wang S., Fontaine G., Behnke C.E., Mende-Mueller L.M., Schappert K., Lee C., Gibson K.M., Miziorko H.M. J. Biol. Chem. 268:4376-4381(1993).

[2] Hruz P.W., Narasimhan C., Miziorko H.M. Biochemistry 31:6842-6847(1992).

Alpha-isopropylmalate and homocitrate synthases signatures (HMGL2)

The following enzymes have been shown [1] to be functionally as well as evolutionary related: - Alpha-isopropylmalate synthase (EC 4.1.3.12) which catalyzes the first step in the biosynthesis of leucine, the condensation of acetyl-CoA and alpha- ketoisovalerate to form 2-isopropylmalate synthase. - Homocitrate synthase (EC 4.1.3.21) (gene nifV) which is involved in the biosynthesis of the iron-molybdenum cofactor of nitrogenase and catalyzes the condensation of acetyl-CoA and alpha-ketoglutarate into homocitrate. - Soybean late nodulin 56. - Methanococcus jannaschii hypothetical proteins MJ0503, MJ1195 and MJ1392. Two conserved regions were selected as signature patterns for these enzymes. The first region is located in the N-terminal section while the second region is located in the central section and contains two conserved histidine residues which could be implicated in the catalytic mechanism.

Consensus pattern: L-R-[DE]-G-x-Q-x(10)-K-

Consensus pattern: [LIVMFYW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI]-

[1] Wang S.-Z., Dean D.R., Chen J.-S., Johnson J.L. J. Bacteriol. 173:3041-3046(1991).

278. (HMG CoA synt) Hydroxymethylglutaryl-coenzyme A synthase active site
Hydroxymethylglutaryl-coenzyme A synthase (EC 4.1.3.5) (HMG-CoA synthase) catalyzes the condensation of acetyl-CoA with acetoacetyl-CoA to produce HMG- CoA and CoA [1]. In vertebrates there are two isozymes located in different subcellular compartments: a cytosolic form which is the starting point of the mevalonate pathway which leads to cholesterol and other sterolic and isoprenoid compounds and a mitochondrial form responsible for ketone body biosynthesis. HMG-CoA is also found in other eukaryotes such as insect, plants and fungi. A cysteine is known to act as the catalytic nucleophile in the first step of the reaction,

the acetylation of the enzyme by acetyl-CoA. The conserved region was used around this active site residue as a signature pattern.

Consensus pattern: N-x-[DN]-[IV]-E-G-[IV]-D-x(2)-N-A-C-[FY]-x-G [C is the active site residue]-

[1] Rokosz L.L., Boulton D.A., Butkiewicz E.A., Sanyal G., Cueto M.A., Lachance P.A., Hermes J.D. Arch. Biochem. Biophys. 312:1-13(1994).

279. HMG (high mobility group) box

280. HSF-type DNA-binding domain signature

Heat shock factor (HSF) is a DNA-binding protein that specifically binds heat shock promoter elements (HSE). HSE is a palindromic element rich with repetitive purine and pyrimidine motifs: 5'-nGAAnnTTCnnGAAnnTTCn-3'. HSF is expressed at normal temperatures but is activated by heat shock or chemical stressors [1,2]. The sequences of HSF from various species show extensive similarity in a region of about 90 amino acids, which has been shown [3] to bind DNA. Some other proteins also contain a HSF domain, these are:

- Yeast SFL1, a protein involved in cell surface assembly and regulation of the gene related to flocculation (asexual cell aggregation) [4].
- Yeast transcription factor SKN7 (or BRY1 or POS9), which binds to the promoter elements SCB and MCB essential for the control of G1 cyclins expression [5].
- Yeast MGA1.
- Yeast hypothetical protein YJR147w.

A pattern from the most conserved part of the HSF DNA-binding domain was derived, its central region.

Consensus pattern: L-x(3)-[FY]-K-H-x-N-x-[STAN]-S-F-[LIVM]-R-Q-L-[NH]-x-Y-x-[FYW]-[RKH]-K-[LIVM]-

[1] Sorger P.K. *Cell* 65:363-366(1991).

[2] Mager W.H., Moradas Ferreira P. *Biochem. J.* 290:1-13(1993).

[3] Vuister G.W., Kim S.-J., Orosz A., Marquardt J., Wu C., Bax A. *Nat. Struct. Biol.* 1:605-613(1994).

- [4] Fujita A., Kikuchi Y., Kuhara S., Misumi Y., Matsumoto S., Kobayashi H. *Gene* 85:321-328(1989).
- [5] Morgan B.A., Bouquin N., Merrill G.F., Johnston L.H. *EMBO J.* 14:5679-5689(1995).

5

281. Heat shock hsp20 proteins family profile

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by inducing the synthesis of proteins collectively known as heat-shock proteins (hsp) [1].

10

Amongst them is a family of proteins with an average molecular weight of 20 Kd, known as the hsp20 proteins [2 to 5]. These seem to act as chaperones that can protect other proteins against heat-induced denaturation and aggregation. Hsp20 proteins seem to form large heterooligomeric aggregates; their family is currently composed of the following members: - Vertebrate heat shock protein hsp27 (hsp25), induced by a variety of environmental stresses. - *Drosophila* heat shock proteins hsp22, hsp23, hsp26, hsp27, hsp67BA and BC. -

15

Caenorhabditis elegans hsp16 multigene family. - Fungal HSP26 (budding yeast) and hsp30 (*Neurospora crassa* and *Aspergillus Nidulans*). - Plant small hsp's. Plants have four classes of hsp20: classes I and II which are cytoplasmic, class III which is chloroplastic and class IV which is found in the endomembrane. - Alpha-crystallin A and B chains. Alpha-crystallin is an abundant constituent of the eye lens of most vertebrate species. Its main function appears to be to maintain the correct refractive index of the lens. It is also found in other tissues where it seems to act as a chaperone [6]. - *Schistosoma mansoni* major egg antigen p40.

20

Structurally, p40 is built of two tandem hsp20 domains. - A variety of prokaryotic proteins: *ibpA* and *ibpB* from *Escherichia coli*, hsp18 from *Clostridium acetobutylicum*, spore protein SP21 (hspA) from *Stigmatella aurantiaca*, *Mycobacterium leprae* 18 Kd antigen and *Mycobacterium tuberculosis* 14 Kd antigen. - *Methanococcus jannaschii* hypothetical protein MJ0285. Structurally, this family is characterized by the presence of a conserved C-terminal domain of about 100 residues. The profile developed to detect members of the hsp20 family is based on an alignment of this domain.

25

-Sequences known to belong to this class detected by the profile: ALL.

30

- [1] Lindquist S., Craig E.A. *Annu. Rev. Genet.* 22:631-677(1988).[2] de Jong W.W., Leunissen J.A.M., Voorter C.E.M. *Mol. Biol. Evol.* 10:103-126(1993).[3] Caspers G.J., Leunissen J.A.M., de Jong W.W. *J. Mol. Evol.* 40:238-248(1995).[4] Jaenicke R., Creighton T.E. *Curr. Biol.* 3:234-235(1993).[5] Jakob U., Buchner J. *Trends Biochem. Sci.* 19:205-

211(1994).[6] Groenen P.J.T.A., Merck K.B., de Jong W.W., Bloemendal H. Eur. J. Biochem. 225:1-9(1994).

282. Heat shock hsp70 proteins family signatures

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by the induction of the synthesis of proteins collectively known as heat-shock proteins (hsp) [1]. Amongst them is a family of proteins with an average molecular weight of 70 Kd, known as the hsp70proteins [2,3,4]. In most species, there are many proteins that belong to the hsp70 family. Some of them are expressed under unstressed conditions. Hsp70proteins can be found in different cellular compartments (nuclear, cytosolic, mitochondrial, endoplasmic reticulum, etc.). Some of the hsp70 family proteinsare listed below: - In Escherichia coli and other bacteria, the main hsp70 protein is known as the dnaK protein. A second protein, hscA, has been recently discovered. dnaK is also found in the chloroplast genome of red algae. - In yeast, at least ten hsp70 proteins are known to exist: SSA1 to SSA4, SSB1, SSB2, SSC1, SSD1 (KAR2), SSE1 (MSI3) and SSE2. - In Drosophila, there are at least eight different hsp70 proteins: HSP70, HSP68, and HSC-1 to HSC-6. - In mammals, there are at least eight different proteins: HSPA1 to HSPA6, HSC70, and GRP78 (also known as the immunoglobulin heavy chain binding protein (BiP)). - In the sugar beet yellow virus (SBYV), a hsp70 homolog has been shown [5] to exist. - In archaeobacteria, hsp70 proteins are also present [6]. All proteins belonging to the hsp70 family bind ATP. A variety of functions has been postulated for hsp70 proteins. It now appears [7] that some hsp70proteins play an important role in the transport of proteins across membranes. They also seem to be involved in protein folding and in the assembly/disassembly of protein complexes [8]. Three signature patterns for the hsp70 family of proteins were derived; the first centered on a conserved pentapeptide found in the N-terminal section of these proteins; the two others on conserved regions located in the central part of the sequence.

Consensus pattern: [IV]-D-L-G-T-[ST]-x-[SC] -

Consensus pattern: [LIVMF]-[LIVMFY]-[DN]-[LIVMFS]-G-[GSH]-[GS]-[AST]-x(3)- [ST]-[LIVM]-[LIVMFC]-

Consensus pattern: [LIVMY]-x-[LIVMF]-x-G-G-x-[ST]-x-[LIVM]-P-x-[LIVM]-x-[DEQKRSTA]-

- [1] Lindquist S., Craig E.A. Annu. Rev. Genet. 22:631-677(1988).
 [2] Pelham H.R.B. Cell 46:959-961(1986).
 [3] Pelham H.R.B. *Nature* 332:776-77(1988).[4] Craig E.A. *BioEssays* 11:48-52(1989).
 5 [5] Agranovsky A.A., Boyko V.P., Karasev A.V., Koonin E.V., Dolja V.V. *J. Mol. Biol.*
 217:603-610(1991).
 [6] Gupta R.S., Singh B. *J. Bacteriol.* 174:4594-4605(1992).
 [7] Deshaies R.J., Koch B.D., Schekman R. *Trends Biochem. Sci.* 13:384-388(1988).
 [8] Craig E.A., Gross C.A. *Trends Biochem. Sci.* 16:135-140(1991).

283. Heat shock hsp90 proteins family signature

Prokaryotic and eukaryotic organisms respond to heat shock or other environmental stress by the induction of the synthesis of proteins collectively known as heat-shock proteins (hsp) [1].
 15 Amongst them is a family of proteins, with an average molecular weight of 90 Kd, known as the hsp90proteins. Proteins known to belong to this family are: - *Escherichia coli* and other bacteria heat shock protein α 62.5 (gene htpG). - Vertebrate hsp 90- α (hsp 86) and hsp 90- β (hsp 84). - *Drosophila* hsp 82 (hsp 83). - *Trypanosoma cruzi* hsp 85. - Plants Hsp82 or Hsp83. - Yeast and other fungi HSC82, and HSP82. - The endoplasmic reticulum protein 'endoplasmic' (also known as Erp99 in mouse, GRP94 in hamster, and hsp 108 in chicken).The exact function of hsp90 proteins is not yet known. In higher eukaryotes, hsp90 has been found associated with steroid hormone receptors, with tyrosine kinase oncogene products of several retroviruses, with eIF2 α kinase, and with actin and tubulin. Hsp90 are probable chaperonins that possess ATPase activity [2,3].As a signature pattern for the hsp90
 20 family of proteins, a highly conserved region found in the N-terminal part of these proteins was selected.

Consensus pattern: Y-x-[NQH]-K-[DE]-[IVA]-F-L-R-[ED] -

- 30 [1] Lindquist S., Craig E.A. Annu. Rev. Genet. 22:631-677(1988).
 [2] Nadeau K., Das A., Walsh C.T. *J. Biol. Chem.* 268:1479-1487(1993).
 [3] Jakob U., Buchner J. *Trends Biochem. Sci.* 19:205-211(1994).

284. Helix-turn-helix (HTH3)

This large family of DNA binding helix-turn helix proteins includes Cro
Swiss:P03036 and CI Swiss:P03034.

285. Heme oxygenase signature

Heme oxygenase (EC 1.14.99.3) (HO) [1] is the microsomal enzyme that, in animals, carries out the oxidation of heme, it cleaves the heme ring at the alpha methene bridge to form biliverdin and carbon monoxide. Biliverdin is subsequently converted to bilirubin by biliverdin reductase. In mammals there are three isozymes of heme oxygenase: HO-1 to HO-3. The first two isozymes differ in their tissue expression and their inducibility: HO-1 is highly inducible by its substrate heme and by various non-heme substances, while HO-2 is non-inducible. It has been suggested [2] that HO-2 could be implicated in the production of carbon monoxide in the brain where it is said to act as a neurotransmitter. In the genome of the chloroplast of red algae as well as in cyanobacteria, there is a heme oxygenase (gene pbsA) that is the key enzyme in the synthesis of the chromophoric part of the photosynthetic antennae [3]. An heme oxygenase is also present in the bacteria *Corynebacterium diphtheriae* (gene hmuO), where it is involved in the acquisition of iron from the host heme [4]. There is, in the central section of these enzymes, a well conserved region centered on a histidine residue which is proposed to play a key role in binding the substrate heme at the active center of the enzyme. This region was used as a signature pattern.

Consensus pattern: L-[IV]-A-H-[STACH]-Y-[STV]-[RT]-Y-[LIVM]-G [H binds the heme] -

[1] Maines M.D. FASEB J. 2:2557-2568(1988).

[2] Barinaga M. Science 259:309-309(1993).

[3] Richaud C., Zabulon G. Proc. Natl. Acad. Sci. U.S.A. 94:11736-11741(1997).

[4] Schmitt M.P. J. Bacteriol. 179:838-845(1997).

286. Hepatitis core antigen.

The core antigen of hepatitis viruses possesses a carboxyl terminus rich in arginine. On this basis it was predicted that the core antigen would bind DNA [1]. There is some experimental evidence to support this [2].

[1] Pasek M, Goto T, Gilbert W, Zink B, Schaller H, McKay P, Leadbetter G, Murray K; Nature 1979;282:575-579. [2] Gallina A, Bonelli F, Zentilin L, Rindi G, Muttini M, Milanese G; J Virol 1989;63:4645-4652.

287. Histidine biosynthesis protein

Proteins involved in steps 4 and 6 of the histidine biosynthesis pathway are contained in this family. Histidine is formed by several complex and distinct biochemical reactions catalysed by eight enzymes. The enzymes in this Pfam entry are called His6 and His7 in eukaryotes and HisA and HisF in prokaryotes.

[1] Fani R, Tamburini E, Mori E, Lazcano A, Lio P, Barberio C, Casalone E, Cavalieri D, Perito B, Polsinelli M, Gene 1997;197:9-17. [2] Fani R, Lio P, Chiarelli I, Bazzicalupo M, J Mol Evol 1994;38:489-495.

288. Histone deacetylase family

Histones can be reversibly acetylated on several lysine residues. Regulation of transcription is caused in part by this mechanism. Histone deacetylases catalyse the removal of the acetyl group. Histone deacetylases are related to other proteins [1].

Leipe DD, Landsman D, Nucleic Acids Res 1997;25:3693-3697.

289. Histidinol dehydrogenase signature

Histidinol dehydrogenase (EC 1.1.1.23) (HDH) catalyzes the terminal step in the biosynthesis of histidine in bacteria, fungi, and plants, the four-electron oxidation of L-histidinol to histidine. In bacteria HDH is a single chain polypeptide; in fungi it is the C-terminal domain of a multifunctional enzyme which catalyzes three different steps of histidine biosynthesis;

and in plants it is expressed as nuclear encoded protein precursor which is exported to the chloroplast [1]. As a signature pattern a highly conserved region located in the central part of HDH was selected. This region does not correspond to the part of the enzyme that, in most, but not all HDH sequences contains a cysteine residue which, in *Salmonella typhimurium*, has been said [2] to be important for the catalytic activity of the enzyme.

Consensus pattern: I-D-x(2)-A-G-P-[ST]-E-[LIVS]-[LIVMA](3)-[AC]-x(3)-A-x(4)-[LIVM]-[AV]-[SACL]-[DE]-[LIVMFC]-[LIVM]-[SA]-x(2)-E-H-

[1] Nagai A., Ward E., Beck J., Tada S., Chang J.-Y., Scheidegger A., Ryals J. Proc. Natl. Acad. Sci. U.S.A. 88:4133-4137(1991).

[2] Grubmeyer C.T., Gray W.R. Biochemistry 25:4778-4784(1986).

290. Homoserine dehydrogenase signature

Homoserine dehydrogenase (EC 1.1.1.3) (HDh) [1,2] catalyzes NAD-dependent reduction of aspartate beta-semialdehyde into homoserine. This reaction is the third step in a pathway leading from aspartate to homoserine. The latter participates in the biosynthesis of threonine and then isoleucine as well as in that of methionine. HDh is found either as a single chain protein as in some bacteria and yeast, or as a bifunctional enzyme consisting of an N-terminal aspartokinase domain and a C-terminal HDh domain as in bacteria such as *Escherichia coli* and in plants. As a signature pattern, the best conserved region of Hdh has been selected. This is a segment of 23 to 24 residues located in the central section and that contains two conserved aspartate residues.

Consensus pattern: A-x(3)-G-[LIVMFY]-[STAG]-x(2,3)-[DNS]-P-x(2)-D-[LIVM]-x-G- x-D-x(3)-K-

[1] Thomas D., Barbey R., Surdin-Kerjan Y. FEBS Lett. 323:289-293(1993).

[2] Cami B., Clepet C., Patte J.-C. Biochimie 75:487-495(1993).

291. haloacid dehalogenase-like hydrolase

This family is structurally different from the alpha/ beta hydrolase family (*abhydrolase*). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment, between residues 10 and 96 of [Swiss:P24069](#). The rest of the fold is composed of the core alpha/beta domain.

[1] Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara T, Esaki N, Soda K, J Biol Chem 1996; 271:20322-20330.

292. DEAD and DEAH box families ATP-dependent helicases signatures (*helicase_C*)

A number of eukaryotic and prokaryotic proteins have been characterized [1,2,3] on the basis of their structural similarity. They all seem to be involved in ATP-dependent, nucleic-acid unwinding. Proteins currently known to belong to this family are: - Initiation factor eIF-4A. Found in eukaryotes, this protein is a subunit of a high molecular weight complex involved in 5'cap recognition and the binding of mRNA to ribosomes. It is an ATP-dependent RNA-helicase. - PRP5 and PRP28. These yeast proteins are involved in various ATP-requiring steps of the pre-mRNA splicing process. - P110, a mouse protein expressed specifically during spermatogenesis. - An3, a *Xenopus* putative RNA helicase, closely related to P110. - SPP81/DED1 and DBP1, two yeast proteins probably involved in pre-mRNA splicing and related to P110. - *Caenorhabditis elegans* helicase *glh-1*. - MSS116, a yeast protein required for mitochondrial splicing. - SPB4, a yeast protein involved in the maturation of 25S ribosomal RNA. - p68, a human nuclear antigen. p68 has ATPase and DNA-helicase activities in vitro. It is involved in cell growth and division. - Rm62 (p62), a *Drosophila* putative RNA helicase related to p68. - DBP2, a yeast protein related to p68. - DHH1, a yeast protein. - DRS1, a yeast protein involved in ribosome assembly. - MAK5, a yeast protein involved in maintenance of dsRNA killer plasmid. - ROK1, a yeast protein. - ste13, a fission yeast protein. - Vasa, a *Drosophila* protein important for oocyte formation and specification of embryonic posterior structures. - Me31B, a *Drosophila* maternally expressed protein of unknown function. - *dbpA*, an *Escherichia coli* putative RNA helicase. - *deaD*, an *Escherichia coli* putative RNA helicase which can suppress a mutation in the *rpsB* gene for ribosomal protein S2. - *rhIB*, an *Escherichia coli* putative RNA helicase. - *rhIE*, an *Escherichia coli* putative RNA helicase. - *srmB*, an *Escherichia coli* protein that shows RNA-dependent ATPase activity. It probably interacts with 23S ribosomal RNA. - *Caenorhabditis elegans*

hypothetical proteins T26G10.1, ZK512.2 and ZK686.2. - Yeast hypothetical protein YHR065c. - Yeast hypothetical protein YHR169w. - Fission yeast hypothetical protein SpAC31A2.07c. - *Bacillus subtilis* hypothetical protein yxiN. All these proteins share a number of conserved sequence motifs. Some of them are specific to this family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' [4,E1]. One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins [3,5,6,E1]. Proteins currently known to belong to this subfamily are: - PRP2, PRP16, PRP22 and PRP43. These yeast proteins are all involved in various ATP-requiring steps of the pre-mRNA splicing process. - Fission yeast prh1, which may be involved in pre-mRNA splicing. - Male-less (mle), a *Drosophila* protein required in males, for dosage compensation of X chromosome linked genes. - RAD3 from yeast. RAD3 is a DNA helicase involved in excision repair of DNA damaged by UV light, bulky adducts or cross-linking agents. Fission yeast rad15 (rhp3) and mammalian DNA excision repair protein XPD (ERCC-2) are the homologs of RAD3. - Yeast CHL1 (or CTF1), which is important for chromosome transmission and normal cell cycle progression in G(2)/M. - Yeast TPS1. - Yeast hypothetical protein YKL078w. - *Caenorhabditis elegans* hypothetical proteins C06E1.10 and K03H1.2. - Poxviruses' early transcription factor 70 Kd subunit which acts with RNA polymerase to initiate transcription from early gene promoters. - I8, a putative vaccinia virus helicase. - hrpA, an *Escherichia coli* putative RNA helicase. Signature patterns were developed for both subfamilies.

Consensus pattern: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]-

Consensus pattern: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR] -

Note: proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop) (see the relevant entry <PDOC00017

[1] Schmid S.R., Linder P. Mol. Microbiol. 6:283-292(1992).

[2] Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K., Schnier J., Slonimski P.P. Nature 337:121-122(1989).

[3] Wassarman D.A., Steitz J.A. Nature 349:463-464(1991).

[4] Hodgman T.C. Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[5] Harosh I., Deschavanne P. *Nucleic Acids Res.* 19:6331-6331(1991).

[6] Koonin E.V., Senkevich T.G. *J. Gen. Virol.* 73:989-993(1992).

5 293. Heme-binding domain in cytochrome b5 and oxidoreductases (heme_1)

Cytochrome b5 is a membrane-bound hemo protein which acts as an electron carrier for several membrane-bound oxygenases [1]. There are two homologous forms of b5, one found in microsomes and one found in the outer membrane of mitochondria. Two conserved histidine residues serve as axial ligands for the heme group. The structure of a number of
10 oxidoreductases consists of the juxtaposition of a heme-binding domain homologous to that of b5 and either a flavodehydrogenase or a molybdopterin domain. These enzymes are:

- Lactate dehydrogenase (EC 1.1.2.3) [2], an enzyme that consists of a flavodehydrogenase domain and a heme-binding domain called cytochrome b2.
- Nitrate reductase (EC 1.6.6.1), a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria [3,4]. Consists of a molybdopterin domain (see <PDOC00484>), a heme-binding domain called cytochrome b557, as
15 well as a cytochrome reductase domain.
- Sulfite oxidase (EC 1.8.3.1) [5], which catalyzes the terminal reaction in the oxidative degradation of sulfur-containing amino acids. Also consists of a
20 molybdopterin domain and a heme-binding domain.

This family of proteins also includes:

- TU-36B, a *Drosophila* muscle protein of unknown function [6].
- Fission yeast hypothetical protein SpAC1F12.10c.
- Yeast hypothetical protein YMR073c.
- Yeast hypothetical protein YMR272c.

A segment was used which includes the first of the two histidine heme ligands, as a signature pattern for the heme-binding domain of cytochrome b5 family.

Consensus pattern: [FY]-[LIVMK]-x(2)-H-P-[GA]-G [H is a heme axial ligand]-

[1] Ozols J. *Biochim. Biophys. Acta* 997:121-130(1989).

[2] Guiard B. *EMBO J.* 4:3265-3272(1985).

[3] Calza R., Huttner E., Vincentz M., Rouze P., Galangau F., Vaucheret H., Cherel I., Meyer C., Kronenberger J., Caboche M. *Mol. Gen. Genet.* 209:552-562(1987).

[4] Crawford N.M., Smith M., Bellissimo D., Davis R.W. *Proc. Natl. Acad. Sci. U.S.A.* 85:5006-5010(1988).

5 [5] Guiard B., Lederer F. *Eur. J. Biochem.* 100:441-453(1979).

[6] Levin R.J., Boychuk P.L., Croniger C.M., Kazzaz J.A., Rozek C.E. *Nucleic Acids Res.* 17:6349-6367(1989).

294. Hexapeptide-repeat containing-transferases signature

On the basis of sequence similarity, a number of transferases have been proposed [1,2,3,4] to belong to a single family. These proteins are: - Serine acetyltransferase (EC 2.3.1.30) (SAT)

(gene *cysE*), an enzyme involved in cysteine biosynthesis. - *Azotobacter chroococcum*

nitrogen fixation protein *nifP*. *NifP* is most probably a SAT involved in the optimization of

15 nitrogenase activity. - *Escherichia coli* thiogalactoside acetyltransferase (EC 2.3.1.18) (gene

lacA), an enzyme involved in the biosynthesis of lactose. - UDP-N-acetylglucosamine

acyltransferase (EC 2.3.1.129) (gene *lpxA*), an enzyme involved in the biosynthesis of lipid

A, a phosphorylated glycolipid that anchors the lipopolysaccharide to the outer membrane of the cell. - UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (EC 2.3.1.-) (gene

20 *lpxD* or *firA*), which is also involved in the biosynthesis of lipid A. - Chloramphenicol

acetyltransferase (CAT) (EC 2.3.1.28) from *Agrobacterium tumefaciens*, *Bacillus sphaericus*,

Escherichia coli plasmid IncFII NR79, *Pseudomonas aeruginosa*, *Staphylococcus aureus*

plasmid *pIP630*. These CAT are not evolutionary related to the main family of CAT (see

<PDOC00093>). - Rhizobium nodulation protein *nodL*. *NodL* is an acetyltransferase

25 involved in the O-acetylation of Nod factors. - Bacterial maltose O-acetyltransferase (EC

2.3.1.79). - Bacterial tetrahydridipicolinate N-succinyltransferase (EC 2.3.1.117) (gene

dapD) which catalyzes the fourth step in the biosynthesis of diaminopimelate and lysine from

aspartate semialdehyde. - Bacterial N-acetylglucosamine-1-phosphate uridyltransferase (EC

2.7.7.23) (gene *glmU* or *gcaD* or *tms*), an enzyme involved in peptidoglycan and

30 lipopolysaccharide biosynthesis. - *Staphylococcus aureus* protein *capG* which is involved in

biosynthesis of type 1 capsular polysaccharide. - Yeast hypothetical protein YJL218w, which

is highly similar to *Escherichia coli lacA*. - Fission yeast hypothetical protein

SpAC18B11.09c. - *Methanococcus jannaschii* hypothetical protein MJ1064. These proteins

have been shown [3,4] to contain a repeat structure composed of tandem repeats of a [LIV]-G-x(4) hexapeptide which, in the tertiary structure of IpxA [5], has been shown to form a left-handed parallel beta helix. Our signature pattern is based on a fourfold repeat of this hexapeptide.

Consensus pattern: [LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIVAC]-x-[LIV]-[GAED]-x(2)-[STAVR]-x-[LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIV]-

[1] Downie J.A. Mol. Microbiol. 3:1649-1651(1989).

[2] Parent R., Roy P.H. J. Bacteriol. 174:2891-2897(1992).

[3] Vaara M. FEMS Microbiol. Lett. 97:249-254(1992).

[4] Vuorio R., Haerakonen T., Tolvanen M., Vaara M. FEBS Lett. 337:289-292(1994).

[5] Ractz C.R.H., Roderick S.L. Science 270:997-1000(1995).

295. Hexokinases signature. Hexokinase (EC 2.7.1.1) [1,2] is an important glycolytic enzyme that catalyzes the phosphorylation of keto- and aldohexoses (e.g. glucose, mannose and fructose) using MgATP as the phosphoryl donor. In vertebrates there are four major isoenzymes, commonly referred as types I,II, III and IV. Type IV hexokinase, which is often incorrectly designated glucokinase [3], is only expressed in liver and pancreatic beta-cells and plays an important role in modulating insulin secretion; it is a protein of a molecular mass of about 50 Kd. Hexokinases of types I to III, which have low Km values for glucose, have a molecular mass of about 100 Kd. Structurally they consist of a very small N-terminal hydrophobic membrane-binding domain followed by two highly similar domains of 450 residues. The first domain has lost its catalytic activity and has evolved into a regulatory domain. In yeast there are three different isozymes: hexokinase PI (gene HXK1), PII (gene HXKB), and glucokinase (gene GLK1). All three proteins have a molecular mass of about 50 Kd. All these enzymes contain one (or two in the case of types I to III isozymes) strongly conserved region which has been shown [4] to be involved in substrate binding. A pattern from that region has been derived

Consensus pattern: [LIVM]-G-F-[TN]-F-S-[FY]-P-x(5)-[LIVM]-[DNST]-x(3)-[LIVM]-x(2)-W-T-K-x-[LF]-

[1] Middleton R.J. Biochem. Soc. Trans. 18:180-183(1990).[2] Griffin L.D., Gelb B.D., Wheeler D.A., Davison D., Adams V., McCabe E.R. Genomics 11:1014-1024(1991).[3] Cornish-Bowden A., Luz Cardenas M. Trends Biochem. Sci. 16:281-282(1991).[4] Schirch D.M., Wilson J.E. Arch. Biochem. Biophys. 254:385-396(1987).

296. Histone H2A signature (his1)

Histone H2A is one of the four histones, along with H2B, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2A sequences [1,2,E1] as a signature pattern, a conserved region in the N-terminal part of H2A. This region is conserved both in classical S-phase regulated H2A's and in variant histone H2A's which are synthesized throughout the cell cycle.

Consensus pattern: [AC]-G-L-x-F-P-V-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).

[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

297. Histone H4 signature (his2)

Histone H4 is one of the four histones, along with H2A, H2B and H3, which forms the eukaryotic nucleosome core. Along with H3, it plays a central role in nucleosome formation. The sequence of histone H4 has remained almost invariant in more than 2 billion years of evolution [1,E1]. The region used as a signature pattern is a pentapeptide found in positions 14 to 18 of all H4 sequences. It contains a lysine residue which is often acetylated [2] and a histidine residue which is implicated in DNA-binding [3].

Consensus pattern: G-A-K-R-H-

[1] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

[2] Doenecke D., Gallwitz D. Mol. Cell. Biochem. 44:113-128(1982).

[3] Ebralidse K.K., Grachev S.A., Mirzabekov A.D. Nature 331:365-367(1988).

Histone H3 signatures (his3)

Histone H3 is one of the four histones, along with H2A, H2B and H4, which forms the eukaryotic nucleosome core. It is a highly conserved protein of 135 amino acid residues [1,2,E1]. The following proteins have been found to contain a C-terminal H3-like domain: -

5 Mammalian centromeric protein CENP-A [3]. Could act as a core histone necessary for the assembly of centromeres. - Yeast chromatin-associated protein CSE4 [4]. - *Caenorhabditis elegans* chromosome III encodes two highly related proteins (F54C8.2 and F58A4.3) whose C-terminal section is evolutionary related to the last 100 residues of H3. The function of these proteins is not yet known. Two signature patterns were developed, The first one corresponds

10 to a perfectly conserved heptapeptide in the N-terminal part of H3. The second one is derived from a conserved region in the central section of H3.

Consensus pattern: K-A-P-R-K-Q-L-

Consensus pattern: P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV]-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).

[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

[3] Sullivan K.F., Hechenberger M., Masri K. J. Cell Biol. 127:581-592(1994).

[4] Stoler S., Keith K.C., Curnick K.E., Fitzgerald-Hayes M. Genes Dev. 9:573-586(1995).

Histone H2B signature (his4)

Histone H2B is one of the four histones, along with H2A, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2B sequences [1,2,E1], a conserved region was selected in the C-terminal part of H2B.

Consensus pattern: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L- x-[KR]-H-A-[LIVM]-[STA]-E-G-

[1] Wells D.E., Brown D. Nucleic Acids Res. 19:2173-2188(1991).

[2] Thatcher T.H., Gorovsky M.A. Nucleic Acids Res. 22:174-179(1994).

297. 'Homeobox' domain signature and profile (home1)

The 'homeobox' is a protein domain of 60 amino acids [1 to 5,E1] first identified in a number of *Drosophila* homeotic and segmentation proteins. It has since been found to be extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Some of the proteins which contain a homeobox domain play an important role in development. Most of these proteins are known to be sequence specific DNA-binding transcription factors. The homeobox domain has also been found to be very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion. A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHHtttHHHHHHHHHxxxxxxxxx ||||| 1
10 20 30 40 50 60 The pattern to detect homeobox sequences that was developed is 24 residues long and spans positions 34 to 57 of the homeobox domain.

Consensus pattern: [LIVMFYGY]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-[LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)- [RKNAIMW] -

[1] Gehring W.J. (In) Guidebook to the homeobox genes, Duboule D., Ed., pp1-10, Oxford University Press, Oxford, (1994).

[2] Buerklin T.R. (In) Guidebook to the homeobox genes, Duboule D., Ed., pp25-72, Oxford University Press, Oxford, (1994).

[3] Gehring W.J. Trends Biochem. Sci. 17:277-280(1992).

[4] Gehring W.J., Hiromi Y. Annu. Rev. Genet. 20:147-173(1986).

[5] Schofield P.N. Trends Neurosci. 10:3-6(1987).

'Homeobox' antennapedia-type protein signature (home2)

The homeotic Hox proteins are sequence-specific transcription factors. They are part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior (A-P) axis [1]. The hox proteins contain a 'homeobox' domain. In *Drosophila* and other insects, there are eight different Hox genes that are encoded in two gene complexes, ANT-C and BX-C. In vertebrates there are 38 genes organized in four complexes. In six of the eight *Drosophila* Hox genes the homeobox domain is highly similar and a

conserved hexapeptide is found five to sixteen amino acids upstream of the homeobox domain. The six *Drosophila* proteins that belong to this group are antennapedia (Antp), abdominal-A (abd-A), deformed (Dfd), proboscipedia (pb), sex combs reduced (scr) and ultrabithorax (ubx) and are collectively known as the 'antennapedia' subfamily. In vertebrates the corresponding Hox genes are known [2] as Hox-A2, A3, A4, A5, A6, A7, Hox-B1, B2, B3, B4, B5, B6, B7, B8, Hox-C4, C5, C6, C8, Hox-D1, D3, D4 and D8. *Caenorhabditis elegans* lin-39 and mab-5 are also members of the 'antennapedia' subfamily. As a signature pattern for this subfamily of homeobox proteins, the conserved hexapeptide was used.

Consensus pattern: [LIVMFE]-[FY]-P-W-M-[KRQTA]-

[1] McGinnis W., Krumlauf R. Cell 68:283-302(1992).

[2] Scott M.P. Cell 71:551-553(1992).

'Homeobox' engrailed-type protein signature (home3)

Most proteins which contain a 'homeobox' domain can be classified [1,2], on the basis of their sequence characteristics, in three subfamilies: engrailed, antennapedia and paired. Proteins currently known to belong to the engrailed subfamily are: - *Drosophila* segmentation polarity protein engrailed (en) which specifies the body segmentation pattern and is required for the development of the central nervous system. - *Drosophila* invected protein (inv). - Silk moth proteins engrailed and invected, which may be involved in the compartmentalization of the silk gland. - Honeybee E30 and E60. - Grasshopper (*Schistocerca americana*) G-En. - Mammalian and birds En-1 and En-2. - Zebrafish Eng-1, -2 and -3. - Sea urchin (*Tripneustes gratilla*) SU-HB-en. - Leech (*Helobdella triserialis*) Ht-En. - *Caenorhabditis elegans* ceh-16. Engrailed homeobox proteins are characterized by the presence of a conserved region of some 20 amino-acid residues located at the C-terminal of the 'homeobox' domain. As a signature pattern for this subfamily of proteins, a stretch of eight perfectly conserved residues in this region was used.

Consensus pattern: L-M-A-[EQ]-G-L-Y-N-

[1] Scott M.P., Tamkun J.W., Hartzell G.W. III *Biochim. Biophys. Acta* 989:25-48(1989).

[2] Gehring W.J. *Science* 236:1245-1252(1987).

298. Isocitrate lyase signature (ICL)

Isocitrate lyase (EC 4.1.3.1) [1,2] is an enzyme that catalyzes the conversion of isocitrate to succinate and glyoxylate. This is the first step in the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. A cysteine, a histidine and a glutamate or aspartate have been found to be important for the enzyme's catalytic activity. Only one cysteine residue is conserved between the sequences of the fungal, plant and bacterial enzymes; it is located in the middle of a conserved hexapeptide that can be used as a signature pattern for this type of enzyme.

Consensus pattern: K-[KR]-C-G-H-[LMQ] [C is a putative active site residue]-

[1] Beeching J.R. Protein Seq. Data Anal. 2:463-466(1989).

[2] Atomi H., Ueda M., Hikida M., Hishida T., Teranishi Y., Tanaka A. J. Biochem. 107:262-266(1990).

299. Initiation factor 2 subunit

This family includes initiation factor 2B alpha, beta and delta subunits from eukaryotes, related proteins from archaeobacteria and IF-2 from prokaryotes. Initiation factor 2 binds to Met-tRNA, GTP and the small ribosomal subunit.

[1] Kyrpides NC, Woese CR, Proc Natl Acad Sci U S A 1998;95:3726-3730.

300. Initiation factor 3 signature

Initiation factor 3 (IF-3) (gene infC) [1] is one of the three factors required for the initiation of protein biosynthesis in bacteria. IF-3 is thought to function as a fidelity factor during the assembly of the ternary initiation complex which consist of the 30S ribosomal subunit, the initiator tRNA and the messenger RNA. IF-3 binds to the 30S ribosomal subunit; it is a basic protein of 141 to 212 residues. The chloroplast initiation factor IF-3(chl) is a protein that enhances the poly(A,U,G)-dependent binding of the initiator tRNA to chloroplast ribosomal 30S subunits. In its mature form it is a protein of about 400 residues whose central

section is evolutionary related to the sequence of bacterial IF-3 [2]. As a signature pattern a highly conserved region was selected located in the central section of bacterial IF-3 and of IF-3(chl).

- 5 Consensus pattern: [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-[LIVMFYS]-x-[FY]-
[DEQTH]-x(2)-[KRQ]-

[1] Liveris D., Schwartz J.J., Geertman R., Schwartz I. FEMS Microbiol. Lett. 112:211-216(1993).

- 10 [2] Lin Q., Ma L., Burkhardt W., Spremulli L.L. J. Biol. Chem. 269:9436-9444(1994).

301. Imidazoleglycerol-phosphate dehydratase signatures (IGPD)

Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) is the enzyme that catalyzes the seventh step in the biosynthesis of histidine in bacteria, fungi and plants. In most organisms it is a monofunctional protein of about 22 to 29 Kd. In some bacteria such as *Escherichia coli* it is the C-terminal domain of a bifunctional protein that include a histidinol-phosphatase domain [1]. Two signature patterns were developed that each include two consecutive histidine residues.

Consensus pattern: [LIVMY]-[DE]-x-H-H-x(2)-E-x(2)-[GCA]-[LIVM]-[STAC]-[LIVM]-

Consensus pattern: G-x-[DN]-x-H-H-x(2)-E-[STAGC]-x-[FY]-K -

[1] Carlomagno M.S., Chiariotti L., Alifano P., Nappo A.G., Bruni C.B. J. Mol. Biol.

- 25 203:585-606(1988).

302. Indole-3-glycerol phosphate synthase signature (IGPS)

Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS) catalyzes the fourth step in the biosynthesis of tryptophan: the ring closure of 1-(2-carboxy-phenylamino)-1-deoxyribulose into indol-3-glycerol-phosphate. In some bacteria, IGPS is a single chain enzyme. In others - such as *Escherichia coli* - it is the N-terminal domain of a bifunctional enzyme that also catalyzes N-(5'-phosphoribosyl)anthranilate isomerase (PRAI) activity, the third step of

303

tryptophan biosynthesis. In fungi, IGPS is the central domain of a trifunctional enzyme that also contains a PRAI C-terminal domain and a glutamine amidotransferase N-terminal domain. The N-terminal section of IGPS contains a highly conserved region which X-ray crystallography studies [1] have shown to be part of the active site cavity. This region was used as a signature pattern for IGPS.

Consensus pattern: [LIVMFY]-[LIVMC]-x-E-[LIVMFYC]-K-[KRSP]-[STAK]-S-P-[ST]-x(3)-[LIVMFYST]-

[1] Wilmanns M., Priestle J.P., Niermann T., Jansonius J.N. J. Mol. Biol. 223:477-507(1992).

303. (IL2) Interleukin 2. 31 members

304. (ILVD EDD) Dihydroxy-acid and 6-phosphogluconate dehydratases. Two dehydratases have been shown [1] to be evolutionary related: - Dihydroxy-acid dehydratase (EC 4.2.1.9) (gene ilvD or ILV3) which catalyzes the fourth step in the biosynthesis of isoleucine and valine, the dehydration of 2,3-dihydroxy-isovaleric acid into alpha-ketoisovaleric acid. - 6-phosphogluconate dehydratase (EC 4.2.1.12) (gene edd) which catalyzes the first step in the Entner-Doudoroff pathway, the dehydration of 6-phospho- D-gluconate into 6-phospho-2-dehydro-3-deoxy-D-gluconate. - Escherichia coli hypothetical protein yjhG. Both enzymes are proteins of about 600 amino acid residues. Two highly conserved regions have been developed as signature patterns. The first pattern is located in the N-terminal part and contains a cysteine that could be involved in the binding of a 2Fe-2S iron-sulfur cluster [2]. The second pattern is located in the C-terminal half.

Consensus pattern: C-D-K-x(2)-P-[GA]-x(3)-[GA] [The C could be a 2Fe-2S ligand]

Consensus pattern: [SA]-L-[LIVM]-T-D-[GA]-R-[LIVMF]-S-[GA]-[GAV]-[ST]-

[1] Egan S.E., Fliege R., Tong S., Shibata A., Wolf R.E. Jr., Conway T. J. *Bacteriol.* 174:4638-4646(1992).[2] Velasco J.A., Cansado J., Pena M.C., Kawakami T., Laborda J., Notario V. *Gene* 137:179-185(1993).

305. IMP dehydrogenase / GMP reductase signature

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2]. GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides. IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

Consensus pattern: [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue]-

[1] Collart F.R., Huberman E. *J. Biol. Chem.* 263:15769-15772(1988).

[2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. *J. Biol. Chem.* 265:5292-5295(1990).

[3] Andrews S.C., Guest J.R. *Biochem. J.* 255:35-43(1988).

306. (IPPe) Inositol polyphosphate phosphatase family, catalytic domain

[1] York JD, Ponder JW, Chen ZW, Mathews FS, Majerus PW; *Biochemistry* 1994;33:13164-13171. [2] Jefferson AB, Auethavekiat V, Pot DA, Williams LT, Majerus PW; *J Biol Chem* 1997;272:5983-5988. [3] Zhang X, Jefferson AB, Auethavekiat V, Majerus PW; *Proc Natl Acad Sci U S A* 1995;92:4853-4856. [4] York JD,

Majerus PW. Proc Natl Acad Sci U S A 1990;87:9548-9552. [5] Neuwald AF, York JD, Majerus PW; FEBS Lett 1991;294:16-18.

5

307. IQ calmodulin-binding motif

[1] Xie X, Harrison DH, Schlichting I, Sweet RM, Kalabokis VN, Szent-Gyorgyi AG, Cohen C; Nature 1994;368:306-312.

10

[2] Rhoads AR, Friedberg F; FASEB J 1997;11:331-340.

308. Inosine-uridine preferring nucleoside hydrolase family signature (IU nuc hydro)

Inosine-uridine preferring nucleoside hydrolase (EC 3.2.2.1) (IU-nucleosidehydrolase or IUNH) is an enzyme first identified in protozoan [1] that catalyzes the hydrolysis of all of the commonly occurring purine and pyrimidine nucleosides into ribose and the associated base, but has a preference for inosine and uridine as substrates. This enzyme is important for these parasitic organisms, which are deficient in de novo synthesis of purines, to salvage the host purine nucleosides. IUNH from Crithidia fasciculata has been sequenced and characterized, it is an homotetrameric enzyme of subunits of 34 Kd. An histidine has been shown to be important for the catalytic mechanism, it acts a proton donor to activate the hypoxanthine leaving group. IUNH is evolutionary related to a number of uncharacterized proteins from various biological sources, notably: - Escherichia coli hypothetical protein yaaF. - Escherichia coli hypothetical protein ybeK. - Escherichia coli hypothetical protein yeiK. - Fission yeast hypothetical protein SpAC17G8.02. - Yeast hypothetical protein YDR400w. - An hypothetical protein from the archaeobacteria Desulfurolobus ambivalens. As a signature pattern for these proteins, a highly conserved region was selected located in the N-terminal extremity. This region contains four conserved aspartates that have been shown [2] to be located in the active site cavity.

30

Consensus pattern: D-x-D-[PT]-[GA]-x-D-D-[TAV]-[VI]-A -

[1] Gopaul D.N., Meyer S.L., Degano M., Sacchettini J.C., Schramm V.L. *Biochemistry* 35:5963-5970(1996).

[2] Degano M., Gopaul D.N., Scapin G., Schramm V.L., Sacchettini J.C. *Biochemistry* 35:5971-5981(1996).

5

309. (Insulinase)

Insulinase family, zinc-binding region signature

(aka Peptidase_M16)

10

A number of proteases dependent on divalent cations for their activity have been shown [1,2] to belong to one family, on the basis of sequence similarity. These enzymes are listed below.

- Insulinase (EC 3.4.24.56) (also known as insulysin or insulin-degrading enzyme or IDE), a cytoplasmic enzyme which seems to be involved in the cellular processing of insulin, glucagon and other small polypeptides.

- *Escherichia coli* protease III (EC 3.4.24.55) (pitirylisin) (gene ptr), a periplasmic enzyme that degrades small peptides.

- Mitochondrial processing peptidase (EC 3.4.24.64) (MPP). This enzyme removes the transit peptide from the precursor form of proteins imported from the cytoplasm across the mitochondrial inner membrane. It is composed of two nonidentical homologous subunits termed alpha and beta. The beta subunit seems to be catalytically active while the alpha subunit has probably lost its activity.

- Nardilysin (EC 3.4.24.61) (N-arginine dibasic convertase or NRD convertase) this mammalian enzyme cleaves peptide substrates on the N-terminus of Arg residues in dibasic stretches.

- *Klebsiella pneumoniae* protein pqqF. This protein is required for the biosynthesis of the coenzyme pyrrolo-quinoline-quinone (PQQ). It is thought to be protease that cleaves peptide bonds in a small peptide (gene pqqA) thus providing the glutamate and tyrosine residues necessary for the synthesis of PQQ.

- Yeast protein AXL1, which is involved in axial budding [3].

- *Eimeria bovis* sporozoite developmental protein.

- *Escherichia coli* hypothetical protein yddC and HI1368, the corresponding *Haemophilus influenzae* protein.
- *Bacillus subtilis* hypothetical protein ymxG.
- *Caenorhabditis elegans* hypothetical proteins C28F5.4 and F56D2.1.

5

It should be noted that in addition to the above enzymes, this family also includes the core proteins I and II of the mitochondrial bc1 complex (also called cytochrome c reductase or complex III), but the situation as to the activity or lack of activity of these subunits is quite complex:

10

- In mammals and yeast, core proteins I and II lack enzymatic activity.
- In *Neurospora crassa* and in potato core protein I is equivalent to the beta subunit of MPP.
- In *Euglena gracilis*, core protein I seems to be active, while subunit II is inactive.

15

These proteins do not share many regions of sequence similarity; the most noticeable is in the N-terminal section. This region includes a conserved histidine followed, two residues later by a glutamate and another histidine. In pitrilysin, it has been shown [4] that this H-x-x-E-H motif is involved in enzyme activity; the two histidines bind zinc and the glutamate is necessary for catalytic activity. Non active members of this family have lost from one to three of these active site residues. We developed a signature pattern that detect active members of this family as well as some inactive members.

20

25

Consensus pattern G-x(8,9)-G-x-[STA]-H-[LIVMFY]-[LIVMC]-[DERN]-[HRKL]-[LMFAT]-x-[LFSTH]-x-[GSTAN]-[GST] [The two H are zinc ligands] [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL active members as well as all MPP alpha subunits and core II subunits. Does not detect inactive core I subunits.

30

Note: these proteins belong to family M16 in the classification of peptidases [5].

[1] Rawlings N.D., Barrett A.J. *Biochem. J.* 275:389-391(1991).

□

- [2] Braun H.-P., Schmitz U.K. Trends Biochem. Sci. 20:171-175(1995).
- [3] Becker A.B., Roth R.A. Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).
- [4] Fujita A., Oka C., Arikawa Y., Katagai T., Tonouchi A., Kuhara S., Misumi Y. Nature 372:567-570(1994).
- [5] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

310. Involucrin repeat

- Eckert RL, Yaffe MB, Crish JF, Murthy S, Rorke EA, Welter JF, *J Invest Dermatol* 1993;100:613-617.

311. Isochorismatase family. This family are hydrolase enzymes.

- Romao MJ, Turk D, Gomis-Ruth FX, Huber R, Schumacher G, Mollering H, Russmann L, J Mol Biol 1992;226:1111-1130.

312. Inositol monophosphatase family signatures (inositol P)

- It has been shown [1] that several proteins share two sequence motifs. Two of these proteins are enzymes of the inositol phosphate second messenger signaling pathway: - Vertebrate and plants inositol monophosphatase (EC 3.1.3.25). - Vertebrate inositol polyphosphate 1-phosphatase (EC 3.1.3.57). The function of the other proteins is not yet clear: - Bacterial protein cysQ. CysQ could help to control the pool of PAPS (3'-phosphoadenoside 5'-phosphosulfate), or be useful in sulfite synthesis. - Escherichia coli protein suhB. Mutations in suhB results in the enhanced synthesis of heat shock sigma factor (htpR). - Neurospora crassa protein Qa-X. Probably involved in quinate metabolism. - Emericella nidulans protein qutG. Probably involved in quinate metabolism. - Yeast protein HAL2/MET22 [2] involved in salt tolerance as well as methionine biosynthesis. - Yeast hypothetical protein YHR046c. - Caenorhabditis elegans hypothetical protein F13G3.5. - A Rhizobium leguminosarum hypothetical protein encoded upstream of the pss gene for exopolysaccharide synthesis. - Methanococcus jannaschii hypothetical protein MJ0109. It is suggested [1] that these proteins may act by enhancing the synthesis or degradation of phosphorylated

messenger molecules. From the X-ray structure of human inositol monophosphatase [3], it seems that some of the conserved residues are involved in binding a metal ion and/or the phosphate group of the substrate.

- 5 Consensus pattern: [FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FY]-x-[HKRNSTY] [The first D and the T bind a metal ion]-
Consensus pattern: [WV]-D-x-[AC]-[GSA]-[GSAPV]-x-[LIVACP]-[LIV]-[LIVAC]-x(3)-[GH]-[GA]-
- 10 [1] Neuwald A.F., York J.D., Majerus P.W. FEBS Lett. 294:16-18(1991).
[2] Glaeser H.-U., Thomas D., Gaxiola R., Montrichard F., Surdin-Kerjan Y., Serrano R. EMBO J. 12:3105-3110(1993).
[3] Bone R., Springer J.P., Atack J.R. Proc. Natl. Acad. Sci. U.S.A. 89:10031-10035(1992).

15 313. Ion transport protein

This family contains Sodium, Potassium, Calcium ion channel This family is 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane. A bacterial structure of the protein is known for the last two helices but is not the Pfam family due to it lacking the first four helices

- 25 314. Isocitrate and isopropylmalate dehydrogenases signature (isodh)
Isocitrate dehydrogenase (IDH) [1,2] is an important enzyme of carbohydrate metabolism which catalyzes the oxidative decarboxylation of isocitrate into alpha-ketoglutarate. IDH is either dependent on NAD⁺ (EC 1.1.1.41) or on NADP⁺(EC 1.1.1.42). In eukaryotes there are at least three isozymes of IDH: two are located in the mitochondrial matrix (one NAD⁺-dependent, the other NADP⁺-dependent), while the third one (also NADP⁺-dependent) is cytoplasmic. In Escherichia coli the activity of a NADP⁺-dependent form of the enzyme is controlled by the phosphorylation of a serine residue; the phosphorylated form of IDH is completely inactivated. 3-isopropylmalate dehydrogenase (EC 1.1.1.85) (IMDH) [3,4]
- 30

catalyzes the third step in the biosynthesis of leucine in bacteria and fungi, the oxidative decarboxylation of 3-isopropylmalate into 2-oxo-4-methylvalerate. Tartrate dehydrogenase (EC [1.1.1.93](#)) [5] catalyzes the reduction of tartrate to oxalloglycolate. These enzymes are evolutionary related [1,3,4,5]. The best conserved region of these enzymes is a glycine-rich stretch of residues located in the C-terminal section. This region was used as a signature pattern.

Consensus pattern: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-[STG]-[LIVMPA]-G-[LIVMF]-

[1] Hurley J.H., Thorsness P.E., Ramalingam V., Helmers N.H., Koshland D.E. Jr., Stroud R.M. Proc. Natl. Acad. Sci. U.S.A. 86:8635-8639(1989).

[2] Cupp J.R., McAlister-Henn L. J. Biol. Chem. 266:22199-22205(1991).

[3] Imada K., Sato M., Tanaka N., Katsube Y., Matsuura Y., Oshima T. J. Mol. Biol. 222:725-738(1991).

[4] Zhang T., Koshland D.E. Jr. Protein Sci. 4:84-92(1995).

[5] Tipton P.A., Beecher B.S. Arch. Biochem. Biophys. 313:15-21(1994).

315. Jacalin-like lectin domain.

Proteins containing this domain are lectins. It is found in 1 to 6 copies in these proteins. The domain is also found in the animal prostatic spermine-binding protein ([Swiss:P15501](#)).

[1] Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Suroolia A, Vijayan M; Nat Struct Biol 1996;3:596-603.

316. KH domain

KH motifs probably bind RNA directly. Auto antibodies to Nova, a KH domain protein, cause paraneoplastic opsoclonus ataxia.

[1] Burd CG, Dreyfuss G, Science 1994;265:615-621.

[2] Musco G, Stier G, Joseph C, Castiglione Morelli MA, Nilges M, Gibson TJ, Pastore A, Cell 1996;85:237-245.

5 317. Kelch motif

The kelch motif was initially discovered in Kelch ([Swiss:Q04652](#)). In this protein there are six copies of the motif. It has been shown that [Swiss:Q04652](#) is related to Galactose Oxidase [1] for which a structure has been solved [2]. The kelch motif forms a beta sheet. Several of these sheets associate to form a beta propeller structure as found in [neur](#).

10 [1] Bork P, Doolittle RF, J Mol Biol 1994;236:1277-1282. [2] Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen, JN, Yadav KD, Knowles PF, Nature 1991;350:87-90.

15 318. Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature

The soybean trypsin inhibitor (Kunitz) family [1] is one of the numerous families of proteinase inhibitors. It comprise plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases as well as some proteins that are probably involved in seed storage. This family is currently known to group the following proteins: - Trypsin inhibitors A, B, C, KTI1, and KTI2 from soybean. - Trypsin inhibitor DE3 from coral beans (*Erythrina* sp.). - Trypsin inhibitor DE5 from sandal bead tree. - Trypsin inhibitors 1A (WTI-1A), 1B (WTI-1B), and 2 (WTI-2) from goa bean. - Trypsin inhibitor from *Acacia confusa*. - Trypsin inhibitor from silk tree. - Chymotrypsin inhibitor 3 (WCI-3) from goa bean. - Cathepsin D inhibitors PDI and NDI 20 from potato [2], which inhibit both cathepsin D (aspartic proteinase) and trypsin. - Alpha-amylase/subtilisin inhibitors from barley and wheat. - Albumin-1 (WBA-1) from goa bean seeds [3]. - Miraculin from *Richadella dulcifica* [4], a sweet taste protein. - Sporamin from sweet potato [5], the major tuberous root protein. - Thiol proteinase inhibitor PCPI 8.3 (P340) 25 from potato tuber [6]. - Wound responsive protein gwin3 from poplar tree [7]. - 21 Kd seed protein from cocoa [8]. All these proteins contain from 170 to 200 amino acid residues and one or two intrachain disulfide bonds. The best conserved region is found in their N-terminal 30 section and is used as a signature pattern

Consensus pattern: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM] -

[1] Laskowski M., Kato I. Annu. Rev. Biochem. 49:593-626(1980).

5 [2] Ritonja A., Krizaj I., Mesko P., Kopitar M., Lucovnik P., Strukelj B., Pungercar J., Buttler D.J., Barrett A.J., Turk V. FEBS Lett. 267:13-15(1990).

[3] Kortt A.A., Strike P.M., de Jersey J. Eur. J. Biochem. 181:403-408(1989).

[4] Theerasilp S., Hitotsuya H., Nakajo S., Nakaja K., Nakamura Y., Kurihara Y. J. Biol. Chem. 264:6655-6659(1989).

10 [5] Hattori T., Yoshida N., Nakamura K. Plant Mol. Biol. 13:563-572(1989).

[6] Krizaj I., Drobnic-Kosorok M., Brzin J., Jerala R., Turk V. FEBS Lett. 333:15-20(1993).

[7] Bradshaw H.D., Hollick J.B., Parsons T.J., Clarke H.R.G., Gordon M.P. Plant Mol. Biol. 14:51-59(1989).

[8] Tai H., McHenry L., Fritz P.J., Furtak D.B. Plant Mol. Biol. 16:913-915(1991).

319. Beta-ketoacyl synthases active site

Beta-ketoacyl-ACP synthase (KAS) [1] is the enzyme that catalyzes the condensation of malonyl-ACP with the growing fatty acid chain. It is found as a component of the following enzymatic systems: - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to different enzymatic activities; beta-ketoacyl synthase is one of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the beta-ketoacyl synthase domain is located in the C-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional chain; the beta-ketoacyl synthase domain is located in the N-terminal section [2]. - The multifunctional 6-methylsalicylic acid synthase (MSAS) from *Penicillium patulum* [3]. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which has a KAS domain in its N-terminal section. - Polyketide antibiotic synthase enzyme systems. Polyketides are secondary metabolites produced by microorganisms and plants from simple fatty acids. KAS is one of the components involved in the biosynthesis of the *Streptomyces* polyketide antibiotics granatacin [4], tetracenomycin C [5] and erythromycin. - *Emericella nidulans* multifunctional protein Wa. Wa is involved in the

biosynthesis of conidial green pigment. Wa is protein of 216 Kd that contains a KAS domain. - Rhizobium nodulation protein node, which probably acts as a beta-ketoacyl synthase in the synthesis of the nodulation Nod factor fatty acyl chain. - Yeast mitochondrial protein CEM1. The condensation reaction is a two step process: the acyl component of an activated acyl primer is transferred to a cysteine residue of the enzyme and is then condensed with an activated malonyl donor with the concomitant release of carbon dioxide. The sequence around the active site cysteine is well conserved and can be used as a signature pattern.

Consensus pattern: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF] [C is the active site residue]

- [1] Kauppinen S., Siggaard-Andersen M., von Wettstein-Knowles P. Carlsberg Res. Commun. 53:357-370(1988).
- [2] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991).
- [3] Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E. Eur. J. Biochem. 192:487-498(1990).
- [4] Bibb M.J., Biro S., Motamedi H., Collins J.F., Hutchinson C.R. EMBO J. 8:2727-2736(1989).
- [5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).

320. Kinesin motor domain signature and profile

Kinesin [1,2,3] is a microtubule-associated force-producing protein that may play a role in organelle transport. Kinesin is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain is composed of three structural domains: a large globular N-terminal domain which is responsible for the motor activity of kinesin (it is known to hydrolyze ATP, to bind and move on microtubules), a central alpha-helical coiled coil domain that mediates the heavy chain dimerization; and a small globular C-terminal domain which interacts with other proteins (such as the kinesin light chains), vesicles and membranous organelles. A number of proteins have been recently found that contain a domain similar to that of the kinesin 'motor'

domain [1,4,E1]: - *Drosophila* claret segregational protein (ncd). Ncd is required for normal chromosomal segregation in meiosis, in females, and in early mitotic divisions of the embryo. The ncd motor activity is directed toward the microtubule's minus end. - *Drosophila* kinesin-like protein (nod). Nod is required for the distributive chromosome segregation of nonexchange chromosomes during meiosis. - Human CENP-E [4]. CENP-E is a protein that associates with kinetochores during chromosome congression, relocates to the spindle midzone at anaphase, and is quantitatively discarded at the end of the cell division. CENP-E is probably an important motor molecule in chromosome movement and/ or spindle elongation. - Human mitotic kinesin-like protein-1 (MKLP-1), a motor protein whose activity is directed toward the microtubule's plus end. - Yeast KAR3 protein, which is essential for yeast nuclear fusion during mating. KAR3 may mediate microtubule sliding during nuclear fusion and possibly mitosis. - Yeast CIN8 and KIP1 proteins which are required for the assembly of the mitotic spindle. Both proteins seem to interact with spindle microtubules to produce an outwardly directed force acting upon the poles. - Fission yeast cut7 protein, which is essential for spindle body duplication during mitotic division. - *Emericella nidulans* bimC, which plays an important role in nuclear division. - *Emericella nidulans* klpA. - *Caenorhabditis elegans* unc-104, which may be required for the transport of substances needed for neuronal cell differentiation. - *Caenorhabditis elegans* osm-3. - *Xenopus* Eg5, which may be involved in mitosis. - *Arabidopsis thaliana* KatA, KatB and katC. - *Chlamydomonas reinhardtii* FLA10/KHP1 and KLP1. Both proteins seem to play a role in the rotation or twisting of the microtubules of the flagella. - *Caenorhabditis elegans* hypothetical protein T09A5.2. The kinesin motor domain is located in the N-terminal part of most of the above proteins, with the exception of KAR3, klpA, and ncd where it is located in the C-terminal section. The kinesin motor domain contains about 330 amino acids. An ATP-binding motif of type A is found near position 80 to 90, the C-terminal half of the domain is involved in microtubule-binding. The signature pattern for that domain is derived from a conserved decapeptide inside the microtubule-binding part.

Consensus pattern: [GSA]-[KRHPSTQVM]-[LIVMF]-x-[LIVMF]-[IVC]-D-L-[AH]-G-[SAN]-E

[1] Bloom G.S., Endow S.A. Protein Prof. 2:1109-1171(1995).

[2] Vallee R.B., Shpetner H.S. Annu. Rev. Biochem. 59:909-932(1990).

[3] Brady S.T. Trends Cell Biol. 5:159-164(1995).

[4] Endow S.A. Trends Biochem. Sci. 16:221-225(1991).[E1]

5 321. Ribosomal protein L15 signature

Ribosomal protein L15 is one of the proteins from the large ribosomal subunit. In *Escherichia coli*, L15 is known to bind the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L15. - Plant chloroplast L15 (nuclear-encoded). - Archaeobacterial L15. - Vertebrate L27a. - *Tetrahymena thermophila* L29. - Fungi L27a (L29, CRP-1, CYH2). L15 is a protein of 144 to 154 amino-acid residues. As a signature pattern, a conserved region was selected in the C-terminal section of these proteins.

Consensus pattern: K-[LIVM](2)-[GASL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-[LIVMFCA]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G

[1] Otake E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

20 322. LBP / BPI / CETP family signature

The following mammalian lipid-binding serum glycoproteins belong to the same family [1,2,3]: - Lipopolysaccharide-binding protein (LBP). LBP binds to the lipid A moiety of bacterial lipopolysaccharides (LPS), a glycolipid present in the outer membrane of all Gram-negative bacteria. The LBP/LPS complex seems to interact with the CD14 receptor and may be responsible for the secretion of alpha-TNF. - Bactericidal permeability-increasing protein (BPI). Like LBP, BPI binds LPS and has a cytotoxic activity on Gram-negative bacteria. - Cholesteryl ester transfer protein (CETP). CETP is involved in the transfer of insoluble cholesteryl esters in reverse cholesterol transport. - Phospholipid transfer protein (PLTP). May play a key role in extracellular phospholipid transport and modulation of HDL particles. These proteins are structurally related and share many regions of sequence similarities. As a signature pattern one of these regions was selected, which is located in the N-terminal section of these proteins; a region which could be involved in the binding to the lipids [2].

Consensus pattern: [PA]-[GA]-[LIVMC]-x(2)-R-[IV]-[ST]-x(3)-L-x(5)-[EQ]-x(4)-[LIVM]-[EQK]-x(8)-P

- [1] Schumann R.R., Leong S.R., Flaggs G.W., Gray P.W., Wright S.D., Mathison J.C., Tobias P.S., Ulevitch R.J. Science 249:1429-1431(1990).
- [2] Gray P.W., Flaggs G., Leong S.R., Gumina R.J., Weiss J., Ooi C.E., Elsbach P. J. Biol. Chem. 264:9505-9509(1989).
- [3] Day J.R., Albers J.J., Lofton-Day C.E., Gilbert T.L., Ching A.F.T., Grant F.J., O'Hara P.J., Marcovina S.M., Adolphson J.L. J. Biol. Chem. 269:9388-9391(1994).

323. LIM domain signature and profile

Recently [1,2] a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues. These proteins are: - *Caenorhabditis elegans* mec-3; a protein required for the differentiation of the set of six touch receptor neurons in this nematode. - *Caenorhabditis elegans* lin-11; a protein required for the asymmetric division of vulval blast cells. - Vertebrate insulin gene enhancer binding protein isl-1. Isl-1 binds to one of the two cis-acting protein-binding domains of the insulin gene. - Vertebrate homeobox proteins lim-1, lim-2 (lim-5) and lim3. - Vertebrate lmx-1, which acts as a transcriptional activator by binding to the FLAT element; a beta-cell-specific transcriptional enhancer found in the insulin gene. - Mammalian LH-2, a transcriptional regulatory protein involved in the control of cell differentiation in developing lymphoid and neural cell types. - *Drosophila* protein apterous, required for the normal development of the wing and halter imaginal discs. - Vertebrate protein kinases LIMK-1 and LIMK-2. - Mammalian rhombotins. Rhombotin 1 (RBTN1 or TTG-1) and rhombotin-2 (RBTN2 or TTG-2) are proteins of about 160 amino acids whose genes are disrupted by chromosomal translocations in T-cell leukemia. - Mammalian and avian cysteine-rich protein (CRP), a 192 amino-acid protein of unknown function. Seems to interact with zyxin. - Mammalian cysteine-rich intestinal protein (CRIP), a small protein which seems to have a role in zinc absorption and may function as an intracellular zinc transport protein. - Vertebrate paxillin, a cytoskeletal focal adhesion protein. - Mouse testin. Mouse testin should not be confused with rat testin which is a thiol protease homolog. - Sunflower pollen specific protein SF3. - Chicken zyxin. Zyxin is a low-abundance adhesion plaque protein which has been shown to interact with CRP. - Yeast protein LRG1

which is involved in sporulation [4]. - Yeast rho-type GTPase activating protein RGA1/DBM1. - *Caenorhabditis elegans* homeobox protein ceh-14. - *Caenorhabditis elegans* homeobox protein unc-97. - Yeast hypothetical protein YKR090w. - *Caenorhabditis elegans* hypothetical proteins C28H8.6. These proteins generally have two tandem copies of a domain, called LIM (for Lin-11 Isl-1 Mec-3) in their N-terminal section. Zyxin and paxillin are exceptions in that they contain respectively three and four LIM domains at their C-terminal extremity. In *apterous*, *isl-1*, *LH-2*, *lin-11*, *lim-1* to *lim-3*, *lmx-1* and *ceh-14* and *mec-3* there is a homeobox domain some 50 to 95 amino acids after the LIM domains. In the LIM domain, there are seven conserved cysteine residues and a histidine. The arrangement followed by these conserved residues is C-x(2)-C-x(16,23)-H-x(2)-[CH]-x(2)-C-x(2)-C-x(16,21)-C-x(2,3)-[CHD]. The LIM domain binds two zinc ions [5]. LIM does not bind DNA, rather it seems to act as an interface for protein-protein interaction. A pattern was developed that spans the first half of the LIM domain.

Consensus pattern: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)-[LIVMF]
[The 5 C's and the H bind zinc]

- [1] Frey G., Kim S.K., Horvitz H.R. *Nature* 344:876-879(1990).
- [2] Baltz R., Evrard J.-L., Domon C., Steinmetz A. *Plant Cell* 4:1465-1466(1992).
- [3] Sanchez-Garcia I., Rabbitts T.H. *Trends Genet.* 10:315-320(1994).
- [4] Mueller A., Xu G., Wells R., Hollenberg C.P., Piepersberg W. *Nucleic Acids Res.* 22:3151-3154(1994).
- [5] Michelsen J.W., Schmeichel K.L., Beckerle M.C., Winge D.R. *Proc. Natl. Acad. Sci. U.S.A.* 90:4404-4408(1993).

324. (LRR) Leucine Rich Repeat

CAUTION: This Pfam may not find all Leucine Rich Repeats in a protein. Leucine Rich Repeats are short sequence motifs present in a number of proteins with diverse functions and cellular locations. These repeats are usually involved in protein-protein interactions. Each Leucine Rich Repeat is composed of a beta-alpha unit. These units form elongated non-globular structures. Leucine Rich Repeats are often flanked by cysteine rich domains.

Number of members: 3017

[1] The leucine-rich repeat: a versatile binding motif. Kobe B, Deisenhofer J; Trends Biochem Sci 1994;19:415-421. [2] Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Kobe B, Deisenhofer J; Nature 1993;366:751-756.

325. Plant lipid transfer protein family signature (LTP)

Plant cells contain proteins, called lipid transfer proteins (LTP) [1,2,3], which are able to facilitate the transfer of phospholipids and other lipids across membranes. These proteins, whose subcellular location is not yet known, could play a major role in membrane biogenesis by conveying phospholipids such as waxes or cutin from their site of biosynthesis to membranes unable to form these lipids. Plant LTP's are proteins of about 9 Kd (90 amino acids) which contain eight conserved cysteine residues all involved in disulfide bridges, as shown in the following schematic representation.

```

      +-----+ +-----+ ||| | *****
15  xCxxxxCxxxxxCxxxxxCxCCCCCCCCxCxxxxxCxx ||| | +-----+ +-----+ +-----+
      -----+

```

'C': conserved cysteine involved in a disulfide bond.

'*': position of the pattern.

20 Consensus pattern: [LIVM]-[PA]-x(2)-C-x-[LIVM]-x-[LIVM]-x-[LIVMFY]-x-[LIVM]-[ST]-x(3)-[DN]-C-x(2)-[LIVM] [The two C's are involved in disulfide bonds]

[1] Wirtz K.W.A. Annu. Rev. Biochem. 60:73-99(1991).

[2] Arondel V., Kader J.C. Experientia 46:579-585(1990).

25 [3] Ohlrogge J.B., Browse J., Somerville C.R. Biochim. Biophys. Acta 1082:1-26(1991).

326. (LAMP) Lysosome-associated membrane glycoproteins signatures

Lysosome-associated membrane glycoproteins (lamp) [1] are integral membrane proteins, specific to lysosomes, and whose exact biological function is not yet clear. Structurally, the lamp proteins consist of two internally homologous lysosome-luminal domains separated by a proline-rich hinge region; at the C-terminal extremity there is a transmembrane region followed by a very short cytoplasmic tail. In each of the duplicated domains, there are two

conserved disulfide bonds. This structure is schematically represented in the figure below. +--

---+ +-----+ +-----+ +-----+ |||||

xCxxxxxCxxxxxxxxxCxxxxxCxxxxxxxxxCxxxxxCxxxxxxxxxCxxxxxCxxxxxxxx <-

-----><Hinge><-----><TM><C> In mammals, there are two

5 closely related types of lamp: lamp-1 and lamp-2. In chicken lamp-1 is known as
LEP100. The macrophage protein CD68 (or macrosialin) [2] is a heavily glycosylated integral
membrane protein whose structure consists of a mucin-like domain followed by a proline-rich
hinge; a single lamp-like domain; a transmembrane region and a short cytoplasmic tail. Two
signature patterns for this family of proteins were developed. The first one is centered on the
10 first conserved cysteine of the duplicated domains. The second corresponds to a region that
includes the extremity of the second domain, the totality of the transmembrane region and the
cytoplasmic tail.

Consensus pattern: [STA]-C-[LIVM]-[LIVMFYW]-A-x-[LIVMFYW]-x(3)-[LIVMFYW]-
15 x(3)-Y [C is involved in a disulfide bond] -

Consensus pattern: C-x(2)-D-x(3,4)-[LIVM](2)-P-[LIVM]-x-[LIVM]-G-x(2)-[LIVM]- x-G-
[LIVM](2)-x-[LIVM](4)-A-[FY]-x-[LIVM]-x(2)-[KR]-[RH]- x(1,2)-[STAG](2)-Y-[EQ] [C
is involved in a disulfide bond]

20 [1] Fukuda M. J. Biol. Chem. 266:21327-21330(1991).

[2] Holness C.L., da Silva R.P., Fawcett J., Gordon S., Simmons D.L. J. Biol. Chem.
268:9661-9666(1993).

25 327. Lipolytic enzymes "G-D-S-L" family, serine active site

Recently [1], a family of lipolytic enzymes has been characterized. This family
currently consist of the following proteins:

- Aeromonas hydrophila lipase/phosphatidylcholine-sterol acyltransferase.
- Xenorhabdus luminescens lipase 1.
- 30 - Vibrio mimicus arylesterase.
- Escherichia coli acyl-coA thioesterase I (gene tesA).
- Vibrio parahaemolyticus thermolabile hemolysin/atypical phospholipase.

- Rabbit phospholipase AdRab-B, an intestinal brush border protein with esterase and phospholipase A/lysophospholipase activity that could be involved in the uptake of dietary lipids. AdRab-B contains four repeats of about 320 amino acids.

- Arabidopsis thaliana and Brassica napus anther-specific proline-rich protein APG.

- 5 - A *Pseudomonas putida* hypothetical protein in trpE-trpG intergenic region. A serine has been identified a part of the active site in the *Aeromonas*, *Vibrio mimicus* and *Escherichia coli* enzymes. It is located in a conserved sequence motif that can be used as a signature pattern for these proteins.

- 10 -Consensus pattern: [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G
[S is the active site residue]

328. (Lipoprotein 4) Prokaryotic membrane lipoprotein lipid attachment site

- 15 In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signalpeptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp). - *Escherichia coli* lipoprotein-28 (gene nlpA). - *Escherichia coli* lipoprotein-34 (gene nlpB). - *Escherichia coli* lipoprotein nlpC. - *Escherichia coli* lipoprotein nlpD. - *Escherichia coli* osmotically inducible lipoprotein B (gene osmB). - *Escherichia coli* osmotically inducible lipoprotein E (gene osmE). - *Escherichia coli* peptidoglycan-associated lipoprotein (gene pal). - *Escherichia coli* rare lipoproteins A and B
- 20 (genes rplA and rplB). - *Escherichia coli* copper homeostasis protein cutF (or nlpE). - *Escherichia coli* plasmids traT proteins. - *Escherichia coli* Col plasmids lysis proteins. - A number of *Bacillus* beta-lactamases. - *Bacillus subtilis* periplasmic oligopeptide-binding protein (gene oppA). - *Borrelia burgdorferi* outer surface proteins A and B (genes ospA and ospB). - *Borrelia hermsii* variable major protein 21 (gene vmp21) and 7 (gene vmp7). -
- 25 *Chlamydia trachomatis* outer membrane protein 3 (gene omp3). - *Fibrobacter succinogenes* endoglucanase cel-3. - *Haemophilus influenzae* proteins Pal and Pcp. - *Klebsiella pullulunase* (gene pulA). - *Klebsiella pullulunase* secretion protein pulS. - *Mycoplasma hyorhinis* protein p37. - *Mycoplasma hyorhinis* variant surface antigens A, B, and C (genes vlpABC). -
- 30

Neisseria outer membrane protein H.8. - *Pseudomonas aeruginosa* lipopeptide (gene *lppL*). - *Pseudomonas solanacearum* endoglucanase *egl*. - *Rhodopseudomonas viridis* reaction center cytochrome subunit (gene *cytC*). - Rickettsia 17 Kd antigen. - *Shigella flexneri* invasion plasmid proteins *mxjI* and *mxjM*. - *Streptococcus pneumoniae* oligopeptide transport protein A (gene *amiA*). - *Treponema pallidum* 34 Kd antigen. - *Treponema pallidum* membrane protein A (gene *tmpA*). - *Vibrio harveyi* chitinase (gene *chb*). - *Yersinia* virulence plasmid protein *yscJ*. - Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification was derived.

Consensus pattern: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).

[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).

[3] von Heijne G. Protein Eng. 2:531-534(1989).

[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

329. (Lipoprotein 5) Prokaryotic membrane lipoprotein lipid attachment site. In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]): - Major outer membrane lipoprotein (murein-lipoproteins) (gene *lpp*). - *Escherichia coli* lipoprotein-28 (gene *nlpA*). - *Escherichia coli* lipoprotein-34 (gene *nlpB*). - *Escherichia coli* lipoprotein *nlpC*. - *Escherichia coli* lipoprotein *nlpD*. - *Escherichia coli* osmotically inducible lipoprotein B (gene *osmB*). - *Escherichia coli* osmotically inducible lipoprotein E (gene *osmE*). - *Escherichia coli* peptidoglycan-associated

lipoprotein (gene pal). - *Escherichia coli* rare lipoproteins A and B (genes rplA and rplB). - *Escherichia coli* copper homeostasis protein cutF (or nlpE). - *Escherichia coli* plasmids traT proteins. - *Escherichia coli* Col plasmids lysis proteins. - A number of *Bacillus* beta-lactamases. - *Bacillus subtilis* periplasmic oligopeptide-binding protein (gene oppA). -

5 *Borrelia burgdorferi* outer surface proteins A and B (genes ospA and ospB). - *Borrelia hermsii* variable major protein 21 (gene vmp21) and 7 (gene vmp7). - *Chlamydia trachomatis* outer membrane protein 3 (gene omp3). - *Fibrobacter succinogenes* endoglucanase cel-3. - *Haemophilus influenzae* proteins Pal and Pcp. - *Klebsiella* pullulanase (gene pulA). - *Klebsiella* pullulanase secretion protein pulS. - *Mycoplasma hyorhinis* protein p37. -

10 *Mycoplasma hyorhinis* variant surface antigens A, B, and C (genes vlp ABC). - *Neisseria* outer membrane protein H.8. - *Pseudomonas aeruginosa* lipopeptide (gene lppL). - *Pseudomonas solanacearum* endoglucanase egl. - *Rhodospseudomonas viridis* reaction center cytochrome subunit (gene cytC). - Rickettsia 17 Kd antigen. - *Shigella flexneri* invasion plasmid proteins mxlJ and mxlM. - *Streptococcus pneumoniae* oligopeptide transport protein

15 A (gene amiA). - *Treponema pallidum* 34 Kd antigen. - *Treponema pallidum* membrane protein A (gene tmpA). - *Vibrio harveyi* chitinase (gene chb). - *Yersinia* virulence plasmid protein yscJ. - Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion). From the precursor sequences of all these proteins, a consensus pattern and a set of

20 rules to identify this type of post-translational modification have been developed.

Consensus pattern: {DERK}{(6)-[LIVFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first

25 seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).[3] von Heijne G. Protein Eng. 2:531-534(1989).[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterheld D., Engelhard M.

30 J. Biol. Chem. 269:14939-14945(1994).

330. (Lum binding) Riboflavin synthase alpha chain family Lum-binding site signature

The following proteins have been shown [1,2] to be structurally and evolutionary related: - Riboflavin synthase alpha chain (RS-alpha) (gene ribC in *Escherichia coli*, ribB in *Bacillus subtilis* and *Photobacterium leiognathi*, RIB5 in yeast). This enzyme synthesizes riboflavin from two moles of 6,7- dimethyl-8-(1'-D-ribityl)lumazine (Lum), a pteridine-derivative. -

5 Photobacterium phosphoreum lumazine protein (LumP) (gene luxL). LumP is a protein that modulates the color of the bioluminescence emission of bacterial luciferase. In the presence of LumP, light emission is shifted to higher energy values (shorter wavelength). LumP binds non-covalently to 6,7-dimethyl-8-(1'-D-ribityl) lumazine. - *Vibrio fischeri* yellow fluorescent protein (YFP) (gene luxY). Like LumP, YFP modulates light emission but towards a longer

10 wavelength. YFP binds non-covalently to FMN. These proteins seem to have evolved from the duplication of a domain of about 100 residues. In its C-terminal section, this domain contains a conserved motif [KR]-V-N-[LI]-E which has been proposed to be the binding site for Lum. RS-alpha which binds two molecules of Lum has two perfect copies of this motif, while LumP which binds one molecule of Lum, has a Glu instead of Lys/Arg in the first

15 position of the second copy of the motif. Similarly, YFP, which binds to one molecule of FMN, also seems to have a potentially dysfunctional binding site by substitution of Gly for Glu in the last position of the first copy of the motif. Our signature pattern includes the Lum-binding motif.

20 Consensus pattern: [LIVMF]-x(5)-G-[STADNQ]-[KREQIYW]-V-N-[LIVM]-E

[1] O'Kane D.J., Woodward B., Lee J., Prasher D.C. Proc. Natl. Acad. Sci. U.S.A. 88:1100-1104(1991).

[2] O'Kane D.J., Prasher D.C. Mol. Microbiol. 6:443-449(1992).

331. Lysyl oxidase putative copper-binding region signature

Lysyl oxidase (LOX) [1] is an extracellular copper-dependent enzyme that catalyzes the oxidative deamination of peptidyl lysine residues in precursors of various collagens and

30 elastins. The deaminated lysines are then able to form aldehyde cross-links. LOX binds a single copper atom which seems to reside within an octahedral coordination complex which includes at least three histidine ligands. Four histidine residues are clustered in a central

region of the enzyme. This region is thought to be involved in cooper-binding and is called the 'copper-talon' [1]. This region was used as a signature pattern.

Consensus pattern: W-E-W-H-S-C-H-Q-H-Y-H

[1] Krebs C.J., Krawetz S.A. Biochim. Biophys. Acta 1202:7-12(1993).

332. Metallo-beta-lactamase superfamily (lactamase_B)

[1] : Neuwald AF, Liu JS, Lipman DJ, Lawrence CE, Nucleic Acids Res 1997;25:1665-1677. [2] Carfi A, Pares S, Duee E, Galleni M, Duez C, Frere JM, Dideberg O, EMBO J 1995;14:4914-4921.

333. L-lactate dehydrogenase active site (ldh1)

L-lactate dehydrogenase (EC 1.1.1.27) (LDH) [1] catalyzes the reversible NAD-dependent interconversion of pyruvate to L-lactate. In vertebrate muscles and in lactic acid bacteria it represents the final step in anaerobic glycolysis. This tetrameric enzyme is present in prokaryotic and eukaryotic organisms. Invertebrates there are three isozymes of LDH: the M form (LDH-A), found predominantly in muscle tissues; the H form (LDH-B), found in heart muscle and the X form (LDH-C), found only in the spermatozoa of mammals and birds. In birds and crocodilian eye lenses, LDH-B serves as a structural protein and is known as epsilon-crystallin [2]. L-2-hydroxyisocaproate dehydrogenase (EC 1.1.1.-) (L-hicDH) [3] catalyzes the reversible and stereospecific interconversion between 2-ketocarboxylic acids and L-2-hydroxy-carboxylic acids. L-hicDH is evolutionary related to LDH's. As a signature for LDH's a region was selected that includes a conserved histidine which is essential to the catalytic mechanism.

Consensus pattern: [LIVMA]-G-[EQ]-H-G-[DN]-[ST] [H is the active site residue] -

[1] Abad-Zapatero C., Griffith J.P., Sussman J.L., Rossmann M.G. J. Mol. Biol. 198:445-467(1987).

[2] Hendriks W., Mulders J.W.M., Bibby M.A., Slingsby C., Bloemendal H., de Jong W.W. Proc. Natl. Acad. Sci. U.S.A. 85:7114-7118(1988).

[3] Lerch H.-P., Frank R., Collins J. Gene 83:263-270(1989).

5 Malate dehydrogenase active site signature (ldh2)

Malate dehydrogenase (EC 1.1.1.37) (MDH) [1,2] catalyzes the interconversion of malate to oxaloacetate utilizing the NAD/NADH cofactor system. The enzyme participates in the citric acid cycle and exists in all aerobic organisms. While prokaryotic organisms contains a single form of MDH, in eukaryotic cells there are two isozymes: one which is located in the
10 mitochondrial matrix and the other in the cytoplasm. Fungi and plants also harbor a glyoxysomal form which functions in the glyoxylate pathway. In plants chloroplast there is an additional NADP-dependent form of MDH (EC 1.1.1.82) which is essential for both the universal C3 photosynthesis (Calvin) cycle and the more specialized C4 cycle. As a signature pattern for this enzyme a region was chosen that includes two residues involved in the
15 catalytic mechanism [3]: an aspartic acid which is involved in a proton relay mechanism, and an arginine which binds the substrate.

Consensus pattern: [LIVM]-T-[TRKMN]-L-D-x(2)-R-[STA]-x(3)-[LIVMFY] [D and R are the active site residues]-

[1] McAlister-Henn L. Trends Biochem. Sci. 13:178-181(1988).

[2] Gietl C. Biochim. Biophys. Acta 1100:217-234(1992).

[3] Birktoft J.J., Rhodes G., Banaszak L.J. Biochemistry 28:6065-6081(1989).

[4] Cendrin F., Chroboczek J., Zaccai G., Eisenberg H., Mevarech M. Biochemistry
25 32:4308-4313(1993).

334. Legume lectins signatures

Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2].

30 These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals). Legume lectins are synthesized as precursor proteins of about 230 to 260

amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal). The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain. Two signature patterns specific to legume lectins have been developed: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain.

Consensus pattern: [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium]-

Consensus pattern: [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST]-

[1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).

[2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).

335. CoA-ligases (ligases- CoA)

This family includes the CoA ligases Succinyl-CoA synthetase alpha: and beta chains, malate CoA ligase and ATP-citrate lyase. Some members of the family utilise ATP others use GTP.

[1] Wolodko WT, Fraser ME, James MN, Bridger WA, J Biol Chem 1994;269:10883-10890.

336. linker histone H1 and H5 family

Linker histone H1 is an essential component of chromatin structure. H1 links nucleosomes into higher order structures Histone H1 is replaced by histone H5 in some cell types.

[1] Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM, Nature 1993;362:219-223.

[3,18]. Others, the outlier lipocalin group, share only one or two of these [3,18]. A signature pattern was built around the first, common to all outlier and kernallipocalins, which occurs near the start of the first beta-strand.

5 Consensus pattern: [DENG]-x-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-
{C}-W-[FYWLRH]-x-[LIVMTA]-

Note: it is suggested, on the basis of similarities of structure, function, and sequence, that this family forms an overall superfamily, called the calycins, with the avidin/streptavidin <PDOC00499> and the cytosolic fatty- acid binding proteins <PDOC00188> families [3,19]

10

[1] Cowan S.W., Newcomer M.E., Jones T.A. Proteins 8:44-61(1990).

[2] Igarashi M., Nagata A., Toh H., Urade H., Hayaishi N. Proc. Natl. Acad. Sci. U.S.A. 89:5376-5380(1992).

[3] Flower D.R., North A.C.T., Attwood T.K. Protein Sci. 2:753-761(1993).

15 [4] Godovac-Zimmermann J. Trends Biochem. Sci. 13:64-66(1988).

[5] Pervaiz S., Brew K. FASEB J. 1:209-214(1987).

[6] Kremer J.M.H., Wilting J., Janssen L.H.M. Pharmacol. Rev. 40:1-47(1989).

[7] Haefliger J.-A., Peitsch M.C., Jenne D., Tschopp J. Mol. Immunol. 28:123-131(1991).

[8] Keen J.N., Caceres I., Eliopoulos E.E., Zagalsky P.F., Findlay J.B.C. Eur. J. Biochem. 197:407-417(1991).

[9] Newcomer M.E. Structure 1:7-18(1993).

[10] Collet C., Joseph R. *Biochim. Biophys. Acta* 1167:219-222(1993).

[11] Kjeldsen L., Johnsen A.H., Sengelov H., Borregaard N. J. Biol. Chem. 268:10425-10432(1993).

25 [12] Peitsch M.C., Boguski M.S. Trends Biochem. Sci. 16:363-363(1991).

[13] Miyawaki A., Matsushita Y.R., Ryo Y., Mikoshiba T. EMBO J. 13:5835-5842(1994).

[14] Kock K., Ahlers C., Schmale H. Eur. J. Biochem. 221:905-916(1994).

[15] Achen M.G., Harms P.J., Thomas T., Richardson S.J., Wettenhall R.E.H., Schreiber G. J. Biol. Chem. 267:23170-23174(1992).

30 [16] Morel L., Dufarre J.-P., Depeiges A. J. Biol. Chem. 268:10274-10281(1993).

[17] Bishop R.E., Penfold S.S., Frost L.S., Holtje J.V., Weiner J.H. J. Biol. Chem. 270:23097-23103(1995).

10

15

30

Note: it is suggested, on the basis of similarities of structure, function, and sequence, that this family forms an overall superfamily, called the calycins, with the lipocalin <PDOC00187> and avidin/streptavidin <PDOC00499> families [6,7].

- [1] Bernier I., Jolles P. *Biochimie* 69:1127-1152(1987).
- [2] Veerkamp J.H., Peeters R.A., Maatman R.G.H.J. *Biochim. Biophys. Acta* 1081:1-24(1991).
- 5 [3] Siegenthaler G., Hotz R., Chatellard-Gruaz D., Didierjean L., Hellman U., Saurat J.-H. *Biochem. J.* 302:363-371(1994).
- [4] Smith A.F., Tsuchida K., Hanneman E., Suzuki T.C., Wells M.A. *J. Biol. Chem.* 267:380-384(1992).
- [5] Moser D., Tendler M., Griffiths G., Klinkert M.-Q. *J. Biol. Chem.* 266:8447-8454(1991).
- 10 [6] Flower D.R., North A.C.T., Attwood T.K. *Protein Sci.* 2:753-761(1993).
- [7] Flower D.R. *FEBS Lett.* 333:99-102(1993).

338. Lipoygenases iron-binding region signatures

15 Lipoygenases (EC 1.13.11.-) are a class of iron-containing dioxygenases which catalyzes the hydroperoxidation of lipids, containing a *cis,cis*-1,4-pentadiene structure. They are common in plants where they may be involved in a number of diverse aspects of plant physiology including growth and development, pest resistance, and senescence or responses to wounding

20 [1]. In mammals a number of lipoygenases isozymes are involved in the metabolism of prostaglandins and leukotrienes [2]. Sequence data is available for the following lipoygenases: - Plant lipoygenases (EC 1.13.11.12). Plants express a variety of cytosolic isozymes as well as what seems [3] to be a chloroplast isozyme. - Mammalian arachidonate 5-lipoygenase (EC 1.13.11.34). - Mammalian arachidonate 12-lipoygenase (EC 1.13.11.31). - Mammalian erythroid cell-specific 15-lipoygenase (EC 1.13.11.33). The iron

25 atom in lipoygenases is bound by four ligands, three of which are histidine residues [4]. Six histidines are conserved in all lipoygenase sequences, five of them are found clustered in a stretch of 40 amino acids. This region contains two of the three zinc-ligands; the other histidines have been shown [5] to be important for the activity of lipoygenases. As signatures for this family of enzymes two patterns in the region of the histidine cluster were

30 selected. The first pattern contains the first three conserved histidines and the second pattern includes the fourth and the fifth.

[1] Vick B.A., Zimmerman D.C. (In) Biochemistry of plants: A comprehensive treatise, Stumpf P.K., Ed., Vol. 9, pp.53-90, Academic Press, New-York, (1987).

[3] Peng Y.L., Shirano Y., Ohta H., Hibino T., Tanaka K., Shibata D. J. Biol. Chem. 269:3755-3761(1994).

[5] Steczko J., Donoho G.P., Clemens J.C., Dixon J.E., Axelrod B. *Biochemistry* 31:4053-4057(1992).

A number of enzymes, belonging to the lyase class, for which fumarate is a substrate have been shown [1,2] to share a short conserved sequence around a methionine which is probably involved in the catalytic activity of this type of enzymes. These enzymes are: - Fumarase (EC

Arginosuccinase (EC [4.3.2.1](#)) (argininosuccinate lyase), which catalyzes the formation of arginine and fumarate from arginosuccinate, the last step in the biosynthesis of arginine. -

Adenylosuccinase (EC 4.3.2.2) (adenylosuccinate lyase) [3], which catalyzes the eight step in the de novo biosynthesis of purines, the formation of 5'-phosphoribosyl-5-amino-4-imidazolecarboxamide and fumarate from 1-(5-phosphoribosyl)-4-(N-succino-carboxamide). That enzyme can also catalyzes the formation of fumarate and AMP from adenylosuccinate.

Pseudomonas putida 3-carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2) (3-carboxymuconate lactonizing enzyme) (gene *pcaB*) [4], an enzyme involved in aromatic acids catabolism

5 Consensus pattern: G-S-x(2)-M-x(2)-K-x-N-

[1] Woods S.A., Shwartzbach S.D., Guest J.R. Biochim. Biophys. Acta 954:14-26(1988).

[2] Woods S.A., Miles J.S., Guest J.R. FEMS Microbiol. Lett. 51:181-186(1988).

[3] Zalkin H., Dixon J.E. Prog. Nucleic Acid Res. Mol. Biol. 42:259-287(1992).

10 [4] Williams S.E., Woolridge E.M., Ransom S.C., Landro J.A., Babbitt P.C., Kozarich J.W. Biochemistry 31:9768-9776(1992).

340. MCM family signature and profile

15 Proteins shown to be required for the initiation of eukaryotic DNA replication share a highly conserved domain of about 210 amino-acid residues [1,2,3]. The latter shows some similarities [4] with that of various other families of DNA-dependent ATPases. Eukaryotes seem to possess a family of six proteins that contain this domain. They were first identified in yeast where most of them have a direct role in the initiation of chromosomal DNA replication

20 by interacting directly with autonomously replicating sequences (ARS). They were thus called 'minichromosome maintenance proteins' with gene symbols prefixed by MCM. These six proteins are: - MCM2, also known as *cdc19* (in *S.pombe*) [E1]. - MCM3, also known as DNA polymerase alpha holoenzyme-associated protein P1, RLF beta subunit or ROA. - MCM4, also known as *CDC54*, *cdc21* (in *S.pombe*) or *dpa* (in *Drosophila*). - MCM5, also

25 known as *CDC46* or *nda4* (in *S.pombe*). - MCM6, also known as *mis5* (in *S.pombe*). - MCM7, also known as *CDC47* or *Prolifera* (in *A.thaliana*). This family is also present in archaebacteria. In *Methanococcus jannaschii* there are four members: MJ0363, MJ0961, MJ1489 and MJEL13. The presence of a putative ATP-binding domain implies that these proteins maybe involved in an ATP-consuming step in the initiation of DNA replication in

30 eukaryotes. As a signature pattern, a perfectly conserved region was selected that represents a special version of the B motif found in ATP-binding proteins.

Consensus pattern: G-[IVT]-[LVAC](2)-[IVT]-D-[DE]-[FL]-[DNST]

- [1] Coxon A., Maundrell K., Kearsey S.E. *Nucleic Acids Res.* 20:5571-5577(1992).
 [2] Hu B., Burkhardt R., Schulte D., Musahl C., Knippers R. *Nucleic Acids Res.* 21:5289-5293(1993).
 5 [3] Tye B.-K. *Trends Cell Biol.* 4:160-166(1994).
 [4] Koonin E.V. *Nucleic Acids Res.* 21:2541-2547(1993).

341. Macrophage migration inhibitory factor family signature (MIF)

10 A protein called macrophage migration inhibitory factor (MIF) [1] seems to exert an important role in host inflammatory responses. It play a pivotal role in the host response to endotoxic shock and appears to serve as a pituitary "stress" hormone that regulates systemic inflammatory responses. MIF is a secreted protein of 115 residues which is not processed from a larger precursor. D-dopachrome tautomerase [2] is a mammalian cytoplasmic enzyme
 15 involved in melanin biosynthesis and that tautomerizes D-dopachrome with concomitant decarboxylation to give 5,6-dihydroxyindole (DHI). It is a protein of 117 residues highly related to MIF. It must be noted that MIF binds glutathione and has been said to be related to glutathione S-transferases. This assertion has been later disproved [3].As a signature pattern for these proteins, a conserved region was selected located in the central section.

20 Consensus pattern: [DE]-P-C-A-x(3)-[LIVM]-x-S-I-G-x-[LIVM]-G-

- [1] Bucala R. *Immunol. Lett.* 43:23-26(1994).
 [2] Odh G., Hindemith A., Rosengren A.-M., Rosengren E., Rorsman H. *Biochem. Biophys.*
 25 *Res. Commun.* 197:619-624(1993).
 [3] Pearson W.R. *Protein Sci.* 3:525-527(1994).

342. MIP family signature

30 Recently the sequence of a number of different proteins, that all seem to be transmembrane channel proteins, has been found to be highly related [1 to 4].These proteins are listed below.
 - Mammalian major intrinsic protein (MIP). MIP is the major component of lens fiber gap junctions. Gap junctions mediate direct exchange of ions and small molecule from one cell to

another. - Mammalian aquaporins [5]. These proteins form water-specific channels that provide the plasma membranes of red cells and kidney proximal and collecting tubules with high permeability to water, thereby permitting water to move in the direction of an osmotic gradient. - Soybean nodulin-26, a major component of the peribacteroid membrane induced during nodulation in legume roots after *Rhizobium* infection. - Plants tonoplast intrinsic proteins (TIP). There are various isoforms of TIP: alpha (seed), gamma, Rt (root), and Wsi (water-stress induced). These proteins may allow the diffusion of water, amino acids and/or peptides from the tonoplast interior to the cytoplasm. - Bacterial glycerol facilitator protein (gene *glpF*), which facilitates the movement of glycerol across the cytoplasmic membrane. - *Salmonella typhimurium* propanediol diffusion facilitator (gene *pduF*). - Yeast FPS1, a glycerol uptake/efflux facilitator protein. - *Drosophila* neurogenic protein 'big brain' (*bib*). This protein may mediate intercellular communication; it may function by allowing the transport of certain molecules(s) and thereby sending a signal for an exodermal cell to become an epidermoblast instead of a neuroblast. - Yeast hypothetical protein YFL054c. - A hypothetical protein from the *pepX* region of *Lactococcus lactis*. The MIP family proteins seem to contain six transmembrane segments. Computer analysis shows that these protein probably arose by a tandem, intragenic duplication event from an ancestral protein that contained three transmembrane segments. As a signature pattern a well conserved region was selected which is located in a probable cytoplasmic loop between the second and third transmembrane regions.

Consensus pattern: [HNQA]-x-N-P-[STA]-[LIVMF]-[ST]-[LIVMF]-[GSTAFY]-

- [1] Reizer J., Reizer A., Saier M.H. Jr. *CRC Crit. Rev. Biochem.* 28:235-257(1993).
- [2] Baker M.E., Saier M.H. Jr. *Cell* 60:185-186(1990).
- [3] Pao G.M., Wu L.-F., Johnson K.D., Hoefte H., Chrispeels M.J., Sweet G., Sandal N.N., Saier M.H. Jr. *Mol. Microbiol.* 5:33-37(1991).
- [4] Wistow G.J., Pisano M.M., Chepelinsky A.B. *Trends Biochem. Sci.* 16:170-171(1991).
- [5] Chrispeels M.J., Agre P. *Trends Biochem. Sci.* 19:421-425(1994).

335

Mandelate racemase (EC 5.1.2.2) (MR) and muconate lactonizing enzyme (EC 5.5.1.1) (MLE) are two bacterial enzymes involved in aromatic acid catabolism. They catalyze mechanistically distinct reactions yet they are related at the level of their primary, quaternary (homooctamer) and tertiary structures [1,2]. A number of other proteins also seem to be evolutionary related to these two enzymes. These are: - The various plasmid-encoded chloromuconate cycloisomerases (EC 5.5.1.7). - *Escherichia coli* protein rspA [3], rspA seems to be involved in the degradation of homoserine lactone (HSL) or of one of its metabolite. - *Escherichia coli* hypothetical protein ycjG. - *Escherichia coli* hypothetical protein yidU. - A hypothetical protein from *Streptomyces ambofaciens* [4]. Two signature patterns have been developed for these enzymes; both contain conserved acidic residues. The second pattern contains an aspartate and a glutamate which are ligands for either a magnesium ion (in MR) or a manganese ion (in MLE).

Consensus pattern: A-x-[SAGCN]-[SAG]-[LIVM]-[DEQ]-x-A-[LA]-x-[DE]-[LIA]-x-[GA]-[KRO]-x(4)-[PSA]-[LIV]-x(2)-L-[LIVMF]-G-

Consensus pattern: [LIVF]-x(2)-D-x-[NH]-x(7)-[ACL]-x(6)-[LIVMF]-x(7)-[LIVM]-E-[DENQ]-P [D and E bind a divalent metal ion]-

[1] Neidhart D.J., Kenyon G.L., Gerlt J.A., Petsko G.A. Nature 347:692-694(1990).

[2] Petsko G.A., Kenyon G.L., Gerlt J.A., Ringe D., Kozarich J.W. Trends Biochem. Sci. 18:372-376(1993).

[3] Huisman G.W., Kolter R. Science 265:537-539(1994).

[4] Schneider D., Aigle B., Leblond P., Simonet J.M., Decaris B. J. Gen. Microbiol. 139:2559-2567(1993).

344. Merozoite Surface Antigen 2 (MSA-2) family

Thomas AW, Carr DA, Carter JM, Lyon JA, Mol Biochem Parasitol 1990;43:211-220.

345. MSP (Major sperm protein) domain.

Major sperm proteins are involved in sperm motility. These proteins oligomerise to form filaments. Partial matches to this domain are also found in other non MSP proteins.

These include [Swiss:P40075](#) and [Swiss:P34593](#).

[1] Bullock TL, Roberts TM, Stewart M, J Mol Biol 1996;263:284-296. [2] King KL,

5 Stewart M, Roberts TM, Seavy M, J Cell Sci 1992;101:847-857.

346. (Matrix) Viral matrix protein. Found in Morbillivirus and paramyxovirus, pneumovirus.

Number of members: 105

10

347. O-methyltransferase (methyltransf)

This family includes a range of O-methyltransferases. These enzymes utilise S-adenosyl methionine.

15 [1] Keller NP, Dischinger HC, Bhatnagar D, Cleveland TE, Ullah AH, Appl Environ Microbiol 1993;59:479-484.

348. Magnesium chelatase, subunit ChII

20 Magnesium-chelatase is a three-component enzyme that catalyses the insertion of Mg²⁺ into protoporphyrin IX. This is the first unique step in the synthesis of (bacterio)chlorophyll. Due to this, it is thought that Mg-chelatase has an important role in channeling intermediates into the (bacterio)chlorophyll branch in response to conditions suitable for photosynthetic growth. ChII and BchD have molecular weight between 38-42 kDa.

25

[1] Walker CJ, Willows RD, Biochem J 1997;327:321-333. [2] Petersen BL, Jensen PE, Gibson LC, Stummann BM, Hunter CN, Henningsen KW, J Bacteriol 1998;180:699-704.

30 349. Plasmid recombination enzyme (Mob_Pre)

With some plasmids, recombination can occur in a site specific manner that is independent of RecA. In such cases, the recombination event requires another protein called

337

Pre. Pre is a plasmid recombination enzyme. This protein is: also known as Mob (conjugative mobilization).

[1] Priebe SD, Lacks SA, J Bacteriol 1989;171:4778-4784.

5

350. Monooxygenase

This family includes diverse enzymes that utilise FAD.

[1] Gatti DL, Palfey BA, Lah MS, Entsch B, Massey V, Ballou DP, Ludwig ML, Science 1994;266:110-114.

10

351. Mov34 family

Members of this family are found in proteasome regulatory subunits, eukaryotic initiation factor 3 (eIF3) subunits and regulators of transcription factors.

[1] Aravind L, Ponting CP, Protein Sci 1998;7:1250-1254. [2] Hershey JW, Asano K, Naranda T, Vornlocher HP, Hanachi P, Merrick WC, Biochimie 1996;78:903-907.

15

352. Myc amino-terminal region (Myc_N_term)

The myc family belongs to the basic helix-loop-helix leucine zipper class of transcription factors, see HLH. Myc forms a heterodimer with Max, and this complex regulates cell growth through direct activation of genes involved in cell replication [2].

[1] Facchini LM, Penn LZ, FASEB J 1998;12:633-651. [2] Grandori C, Eisenman RN, Trends Biochem Sci 1997;22:177-181.

25

353. (Metallothio_2) Metallothionein. Members of this family are metallothioneins. These proteins are cysteine rich proteins that bind to heavy metals. Members of this family appear to be closest to Class II metallothioneins, seed metalthio. Number of members: 55

30

[1] Medline: 98267202. Characterization of gene repertoires at mature stage of citrus fruits through random sequencing and analysis of redundant metallothionein-like genes expressed

during fruit development. Moriguchi T, Kita M, Hisada S, Endo-Inagaki T, Omura M; Gene 1998;211:221-227.

5 354. MAGE family

The MAGE (melanoma antigen-encoding gene) family are expressed in a wide variety of tumors but not in normal cells, with the exception of the male germ cells, placenta, and, possibly, cells of the developing embryo. The cellular function of this family is unknown.

[1] McCurdy DK, Tai LQ, Nguyen J, Wang Z, Yang HM, Udar N, Naiem F, Concannon P, Gatti RA; Mol Genet Metab 1998;63:3-13.

15

20

25

30

355. Malic enzymes signature. Malic enzymes, or malate oxidoreductases, catalyze the oxidative decarboxylation of malate into pyruvate important for a wide range of metabolic pathways. There are three related forms of malic enzyme [1,2,3]: - NAD-dependent malic enzyme (EC 1.1.1.38), which uses preferentially NAD and has the ability to decarboxylate oxaloacetate (OAA). It is found in bacteria and insects. - NAD-dependent malic enzyme (EC 1.1.1.39), which uses preferentially NAD and is unable to decarboxylate OAA. It is found in the mitochondrial matrix of plants and is a heterodimer of highly related subunits. - NADP-dependent malic enzyme (EC 1.1.1.40), which has a preference for NADP and has the ability to decarboxylate OAA. This form has been found in fungi, animals and plants. In mammals, there are two isozymes: one, mitochondrial and the other, cytosolic. Plants also have two isozymes: chloroplastic and cytosolic. There are two other proteins which are closely structurally related to malicenzymes: - Escherichia coli protein sfcA, whose function is not yet known but which could be an NAD or NADP-dependent malic enzyme. - Yeast hypothetical protein YKL029c, a probable malic enzyme. There are three well conserved regions in the enzyme sequences. Two of them seem to be involved in binding NAD or NADP. The significance of the third one, located in the central part of the enzymes, is not yet known. This region has been developed as a signature pattern for these enzymes.

Consensus pattern: F-x-[DV]-D-x(2)-G-T-[GSA]-x-[IV]-x-[LIVMA]-[GAST](2)-[LIVMF](2)-

- 5 [1] Artus N.N., Edwards G.E. FEBS Lett. 182:225-233(1985). [2] Loeber G., Infante A.A., Maurer-Fogy I., Krystek E., Dworkin M.B. J. Biol. Chem. 266:3016-3021(1991). [3] Long J.J., Wang J.-L., Berry J.O. J. Biol. Chem. 269:2827-2833(1994).

10 356. (matrixin)

Matrixins cysteine switch (aka peptidase_M10)

Mammalian extracellular matrix metalloproteinases (EC 3.4.24.-), also known as matrixins [1] (see <PDOC00129>), are zinc-dependent enzymes. They are secreted by cells in an inactive form (zymogen) that differs from the mature enzyme by the presence of an N-terminal propeptide. A highly conserved octapeptide is found two residues downstream of the C-terminal end of the propeptide. This region has been shown to be involved in autoinhibition of matrixins [2,3]; a cysteine within the octapeptide chelates the active site zinc ion, thus inhibiting the enzyme. This region has been called the 'cysteine switch' or 'autoinhibitor region'.

A cysteine switch has been found in the following zinc proteases:

- MMP-1 (EC 3.4.24.7) (interstitial collagenase).
- 25 - MMP-2 (EC 3.4.24.24) (72 Kd gelatinase).
- MMP-3 (EC 3.4.24.17) (stromelysin-1).
- MMP-7 (EC 3.4.24.23) (matrilysin).
- MMP-8 (EC 3.4.24.34) (neutrophil collagenase).
- MMP-9 (EC 3.4.24.35) (92 Kd gelatinase).
- 30 - MMP-10 (EC 3.4.24.22) (stromelysin-2).
- MMP-11 (EC 3.4.24.-) (stromelysin-3).
- MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).
- MMP-13 (EC 3.4.24.-) (collagenase 3).

340

- MMP-14 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 1).
- MMP-15 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 2).
- MMP-16 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 3).
- Sea urchin hatching enzyme (EC 3.4.24.12) (envelysin) [4].
- 5 - Chlamydomonas reinhardtii gamete lytic enzyme (GLE) [5].

Consensus pattern P-R-C-[GN]-x-P-[DR]-[LIVSAPKQ] [C chelates the zinc ion] Sequences known to belong to this class detected by the pattern ALL, except for cat MMP-7 and mouse MMP-11.

10

[1] Woessner J. Jr. FASEB J. 5:2145-2154(1991).

□

[2] Sanchez-Lopez R., Nicholson R., Gesnel M.C., Matrisian L.M., Breathnach R. J. Biol. Chem. 263:11892-11899(1988).

15

[3] Park A.J., Matrisian L.M., Kells A.F., Pearson R., Yuan Z., Navre M. J. Biol. Chem. 266:1584-1590(1991).

[4] Lepage T., Gache C. EMBO J. 9:3003-3012(1990).

[5] Kinoshita T., Fukuzawa H., Shimada T., Saito T., Matsuda Y. Proc. Natl. Acad. Sci. U.S.A. 89:4693-4697(1992).

20

357. Vertebrate metallothioneins signature (metalthio)

Metallothioneins (MT) [1,2,3] are small proteins which bind heavy metals such as zinc, copper, cadmium, nickel, etc., through clusters of thiolate bonds. MT's occur throughout the animal kingdom and are also found in higher plants, fungi and some prokaryotes. On the basis of structural relationships MT's have been subdivided into three classes. Class I includes mammalian MT's as well as MT's from crustacean and molluscs, but with clearly related primary structure. Class II groups together MT's from various species such as sea urchins, fungi, insects and cyanobacteria which display none or only very distant correspondence to class I MT's. Class III MT's are atypical polypeptides containing gamma-glutamylcysteinyl units. Vertebrate class I MT's are proteins of 60 to 68 amino acid residues, 20 of these residues are cysteines that bind to 7 bivalent metal ions. As a signature pattern a region that

25

30

09052002-061600

spans 19 residues and which contains seven of the metal-binding cysteines was chosen, this region is located in the N-terminal section of class-I MT's.

Consensus pattern: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K-

- [1] Hamer D.H. Annu. Rev. Biochem. 55:913-951(1986).
- [2] Kagi J.H.R., Schaffer A. Biochemistry 27:8509-8515(1988).
- [3] Binz P.-A. Thesis, 1996, University of Zurich.

358. Mitochondrial energy transfer proteins signature (mito_carr)

Different types of substrate carrier proteins involved in energy transfer are found in the inner mitochondrial membrane [1 to 5]. These are: - The ADP,ATP carrier protein (AAC) (ADP/ATP translocase) which exports ATP into the cytosol and imports ADP into the mitochondrial matrix. The sequence of AAC has been obtained from various mammalian, plant and fungal species. - The 2-oxoglutarate/malate carrier protein (OGCP), which exports 2-oxoglutarate into the cytosol and imports malate or other dicarboxylic acids into the mitochondrial matrix. This protein plays an important role in several metabolic processes such as the malate/aspartate and the oxoglutarate/isocitrate shuttles. - The phosphate carrier protein, which transports phosphate groups from the cytosol into the mitochondrial matrix. - The brown fat uncoupling protein (UCP) which dissipates oxidative energy into heat by transporting protons from the cytosol into the mitochondrial matrix. - The tricarboxylate transport protein (or citrate transport protein) which is involved in citrate-H⁺/malate exchange. It is important for the bioenergetics of hepatic cells as it provides a carbon source for fatty acid and sterol biosyntheses, and NAD for the glycolytic pathway. - The Grave's disease carrier protein (GDC), a protein of unknown function recognized by IgG in patients with active Grave's disease. - Yeast mitochondrial proteins MRS3 and MRS4. The exact function of these proteins is not known. They suppress a mitochondrial splice defect in the first intron of the COB gene and may act as carriers, exerting their suppressor activity by modulating solute concentrations in the mitochondrion. - Yeast mitochondrial FAD carrier protein (gene FLX1). - Yeast protein ACR1 [6], which seems essential for acetyl-CoA synthetase activity. - Yeast protein PET8. - Yeast protein PMT. - Yeast protein RIM2. - Yeast protein YHM1/SHM1. - Yeast protein YMC1. - Yeast protein YMC2. - Yeast hypothetical

342

proteins YBR291c, YEL006w, YER053c, YFR045w, YHR002w, and YIL006w. -
Caenorhabditis elegans hypothetical protein K11H3.3. Two other proteins have been found to
 belong to this family, yet are not localized in the mitochondrial inner membrane: - Maize
 amyloplast Brittle-1 protein. This protein, found in the endosperm of kernels, could play a
 5 role in amyloplast membrane transport. - *Candida boidinii* peroxisomal membrane protein
 PMP47 [7]. PMP47 is an integral membrane protein of the peroxisome and it may play a role
 as a transporter. These proteins all seem to be evolutionary related. Structurally, they
 consist of three tandem repeats of a domain of approximately one hundred residues. Each of
 these domains contains two transmembrane regions. As a signature pattern, one of the most
 10 conserved regions in the repeated domain was selected, located just after the first
 transmembrane region.

Consensus pattern: P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]-[QGAIVM]-

- [1] Klingenberg M. Trends Biochem. Sci. 15:108-112(1990).
- [2] Walker J.E. Curr. Opin. Struct. Biol. 2:519-526(1992).
- [3] Kuan J., Saier M.H. Jr. CRC Crit. Rev. Biochem. 28:209-233(1993).
- [4] Kuan J., Saier M.H. Jr. Res. Microbiol. 144:671-672(1993).
- [5] Nelson D.R., Lawson J.E., Klingenberg M., Douglas M.G. J. Mol. Biol. 230:1159-
 20 1170(1993).
- [6] Palmieri F. FEBS Lett. 346:48-54(1994).
- [7] Jank B., Habermann B., Schweyen R.J., Link T.A. Trends Biochem. Sci. 18:427-
 428(1993).

359. Prokaryotic molybdopterin oxidoreductases signatures (molybdopterin)

A number of different prokaryotic oxidoreductases that require and bind a molybdopterin
 cofactor have been shown [1,2,3] to share a number of regions of sequence similarity. These
 enzymes are: - *Escherichia coli* respiratory nitrate reductase (EC 1.7.99.4). This enzyme
 30 complex allows the bacteria to use nitrate as an electron acceptor during anaerobic growth.
 The enzyme is composed of three different chains: alpha, beta and gamma. The alpha chain
 (gene narG) is the molybdopterin-binding subunit. *Escherichia coli* encodes for a second,
 closely related, nitrate reductase complex which also contains a molybdopterin-binding alpha

chain (gene narZ). - *Escherichia coli* anaerobic dimethyl sulfoxide reductase (DMSO reductase). DMSO reductase is the terminal reductase during anaerobic growth on various sulfoxide and N-oxide compounds. DMSO reductase is composed of three chains: A, B and C. The A chain (gene dmsA) binds molybdopterin. - *Escherichia coli* biotin sulfoxide reductases (genes bisC and bisZ). This enzyme reduces a spontaneous oxidation product of biotin, BDS, back to biotin. It may serve as a scavenger, allowing the cell to use biotin sulfoxide as a biotin source. - *Methanobacterium formicicum* formate dehydrogenase (EC 1.2.1.2). The alpha chain (gene fdhA) of this dimeric enzyme binds a molybdopterin cofactor. - *Escherichia coli* formate dehydrogenases -H (gene fdhF), -N (gene fdnG) and -O (gene fdoG). These enzymes are responsible for the oxidation of formate to carbon dioxide. In addition to molybdopterin, the alpha (catalytic) subunit also contains an active site, selenocysteine. - *Wolinella succinogenes* polysulfide reductase chain. This enzyme is a component of the phosphorylative electron transport system with polysulfide as the terminal acceptor. It is composed of three chains: A, B and C. The A chain (gene psrA) binds molybdopterin. - *Salmonella typhimurium* thiosulfate reductase (gene phsA). - *Escherichia coli* trimethylamine-N-oxide reductase (EC 1.6.6.9) (gene torA) [4]. - Nitrate reductase (EC 1.7.99.4) from *Klebsiella pneumoniae* (gene nasA), *Alcaligenes eutrophus*, *Escherichia coli*, *Rhodobacter sphaeroides*, *Thiosphaera pantotropha* (gene napA), and *Synechococcus* PCC 7942 (gene narB). These proteins range from 715 amino acids (fdhF) to 1246 amino acids (narZ) in size. Three signature patterns for these enzymes were derived. The first is based on a conserved region in the N-terminal section and contains two cysteine residues perhaps involved in binding the molybdopterin cofactor. It should be noted that this region is not present in bisC. The second pattern is derived from a conserved region located in the central part of these enzymes.

Consensus pattern: [STAN]-x-[CH]-x(2,3)-C-[STAG]-[GSTVMF]-x-C-x-[LIVMFYW]-x-[LIVMA]-x(3,4)-[DENQKHT]-

Consensus pattern: [STA]-x-[STAC](2)-x(2)-[STA]-D-[LIVMY](2)-L-P-x-[STAC](2)-x(2)-E-

Consensus pattern: A-x(3)-[GDT]-I-x-[DNQTK]-x-[DEA]-x-[LIVM]-x-[LIVMC]-x-[NS]-x(2)-[GS]-x(5)-A-x-[LIVM]-[ST]-

- [1] Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray R.C. *Biochim. Biophys. Acta* 1057:157-185(1991).
- [2] Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H. *Mol. Microbiol.* 2:785-795(1988).
- [3] Trieber C.A., Rothery R.A., Weiner J.H. *J. Biol. Chem.* 269:7103-7109(1994).
- 5 [4] Mejean V., Lobbi-Nivol C., Lepelletier M., Giordano G., Chippaux M., Pascal M.-C. *Mol. Microbiol.* 11:1169-1179(1994).

360. Bacterial mutT domain signature

- 10 The bacterial mutT protein is involved in the GO system [1] responsible for removing an oxidatively damaged form of guanine (8-hydroxyguanine or 7,8-dihydro-8-oxoguanine) from DNA and the nucleotide pool. 8-oxo-dGTP is inserted opposite to dA and dC residues of template DNA with almost equal efficiency thus leading to A.T to G.C transversions. MutT specifically degrades 8-oxo-dGTP to the monophosphate with the concomitant release of
- 15 pyrophosphate. MutT is a small protein of about 12 to 15 Kd. It has been shown [2,3] that a region of about 40 amino acid residues, which is found in the N-terminal part of mutT, can also be found in a variety of other prokaryotic, viral, and eukaryotic proteins. These proteins are:
- *Streptomyces pneumoniae* mutX.
 - 20 - A mutT homolog from plasmid pSAM2 of *Streptomyces ambifaciens*.
 - *Bartonella bacilliformis* invasion protein A (gene invA).
 - *Escherichia coli* dATP pyrophosphohydrolase.
 - Protein D250 from African swine fever viruses.
 - Proteins D9 and D10 from a variety of poxviruses.
 - 25 - Mammalian 7,8-dihydro-8-oxoguanine triphosphatase (EC 3.1.6.-) [4].
 - Mammalian diadenosine 5',5'''-P1,P4-tetraphosphate asymmetrical hydrolase (Ap4Aase) (EC 3.6.1.17) [5], which cleaves A-5'-PPPP-5'A to yield AMP and ATP.
 - A protein encoded on the antisense RNA of the basic fibroblast growth factor gene in higher vertebrates.
 - 30 - Yeast protein YSA1.
 - *Escherichia coli* hypothetical protein yfaO.

- *Escherichia coli* hypothetical protein ygdU and HI0901, the corresponding *Haemophilus influenzae* protein.
- *Escherichia coli* hypothetical protein yjaD and HI0432, the corresponding *Haemophilus influenzae* protein.
- 5 - *Escherichia coli* hypothetical protein yrfE.
- *Bacillus subtilis* hypothetical protein yqkG.
- *Bacillus subtilis* hypothetical protein yzgD.
- Yeast hypothetical protein YGL067w.

It is proposed [2] that the conserved domain could be involved in the active center of a family of pyrophosphate-releasing NTPases. As a signature pattern the core region of the domain was selected; it contains four conserved glutamate residues.

Consensus pattern: G-x(5)-E-x(4)-[STAGC]-[LIVMAC]-x-R-E-[LIVMFT]-x-E-E-

- [1] Michaels M.L., Miller J.H. J. Bacteriol. 174:6321-6325(1992).
- [2] Koonin E.V. Nucleic Acids Res. 21:4847-4847(1993).
- [3] Mejean V., Salles C., Bullions M.J., Bessman M.J., Claverys J.-P. Mol. Microbiol. 11:323-330(1994).
- [4] Sakumi K., Furuichi M., Tsuzuki T., Kakuma T., Kawabata S., Maki H., Sekiguchi M. J. Biol. Chem. 268:23524-23530(1993).
- [5] Thorne N.M.H., Hankin S., Wilkinson M.C., Nunez C., Barraclough R., McLennan A.G. Biochem. J. 311:717-721(1995).

361. Myb DNA-binding domain repeat signatures

The retroviral oncogene v-myb, and its cellular counterpart c-myb, encode nuclear DNA-binding proteins that specifically recognize the sequence YAAC(G/T)G [1]. The myb family also includes the following proteins: - *Drosophila* D-myb [2]. - Vertebrate myb-like proteins A-myb and B-myb [3]. - Maize C1 protein, a trans-acting factor which controls the expression of genes involved in anthocyanin biosynthesis. - Maize P protein [4], a trans-acting factor which regulates the biosynthetic pathway of a flavonoid-derived pigment in certain floral tissues. - *Arabidopsis thaliana* protein GL1 [5], required for the initiation of differentiation of leaf hair cells (trichomes). - A number of myb/c1-related proteins in maize and barley, whose roles are not yet known [4]. - Yeast BAS1 [7], a transcriptional activator

for the *HIS4* gene. - Yeast *REB1* [8], which recognizes sites within both the enhancer and the promoter of rRNA transcription, as well as upstream of many genes transcribed by RNA polymerase II. - Fission yeast *cdc5*, a possible transcription factor whose activity is required for cell cycle progression and growth during G2. - Fission yeast *myb1*, which regulates telomere length and function. - Yeast hypothetical protein *YMR213w*. One of the most conserved regions in all of these proteins is a domain of 160 amino acids. It consists of three tandem repeats of 51 to 53 amino acids. In *myb*, this repeat region has been shown [9] to be involved in DNA-binding. The major part of the first repeat is missing in retroviral v-myb sequences and in plant myb-related proteins. Yeast *REB1* differs from the other proteins in this family in having a single myb-like domain. As shown in the following schematic representation, two signature patterns for myb-like domains were developed; the first is located in the N-terminal section, the second spans the C-terminal extremity of the domain.

XXXXXXXXXXWxxxEDXXXXXXXXXXXXXWxxlxxxxxxRXXXXXXXXXWxxxx *****
 ***** : Position of the patterns.

Consensus pattern: W-[ST]-x(2)-E-[DE]-x(2)-[LIV]-

Consensus pattern: W-x(2)-[LI]-[SAG]-x(4,5)-R-x(8)-[YW]-x(3)-[LIVM]-

Note: this pattern detects the three copies of the domain in *myb*, d-*myb*, A-*myb* and B-*myb*; the second of the two complete copies of plant myb-related proteins, and the last two copies of yeast *BAS1*

[1] Biednkapp H., Borgmeyer U., Sippel A.E., Klempnauer K.-H. *Nature* 335:835-837(1988).

[2] Peters C.W.B., Sippel A.E., Vingron M., Klempnauer K.-H. *EMBO J.* 6:3085-3090(1987).

[3] Nomura N., Takahashi M., Matsui M., Ishii S., Date T., Sasamoto S., Ishizaki R. *Nucleic Acids Res.* 16:11075-11090(1988).

[4] Grotewold E., Athma P., Peterson T. *Proc. Natl. Acad. Sci. U.S.A.* 88:4587-4591(1991).

[5] Oppenheimer D.G., Herman P.L., Sivakumaran S., Esch J., Marks M.D. *Cell* 67:483-493(1991).

[6] Marocco A., Wissenbach M., Becker D., Paz-Ares J., Saedler H., Salamini F., Rohde W. *Mol. Gen. Genet.* 216:183-187(1989).

[7] Tice-Baldwin K., Fink G.R., Arndt K.T. *Science* 246:931-935(1989).

[8] Ju Q., Morrow B.E., Warner J.R. Mol. Cell. Biol. 10:5226-5234(1990).

[9] Klempnauer K.-H., Sippel A.E. EMBO J. 6:2719-2725(1987).

- 5 362. NAD-dependent glycerol-3-phosphate dehydrogenase signature
NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3- phosphate. It is a eukaryotic cytosolic homodimeric protein of about 40 Kd. As a signature pattern a glycine-rich region that is probably [1] involved in NAD-binding was selected.

10

Consensus pattern: G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x-[DE]-
G-[LIVM]-x-[LIVMFYW]-G-x-N-

[1] Otto J., Argos P., Rossmann M.G. Eur. J. Biochem. 109:325-330(1980).

15

363. Nucleosome assembly protein (NAP)

It is thought that NAPs may be involved in regulating gene expression as a result of histone accessibility [1].

20

[1] Rodriguez P, Munroe D, Prawitt D, Chu LL, Bric E, Kim J, Reid LH, Davies C, Nakagama H, Loeber R, Winterpacht A, Petrucci MJ, Higgins MJ, Nowak N, Evans G, Shows T, Weissman BE, Zabel B, Housman DE, Pelletier J, Genomics 1997;44:253-265. [2] Schnieders F, Dork T, Arnemann J, Vogel T, Werner M, Schmidtke J; Hum Mol Genet 1996;5:1801-1807.

25

364. NB-ARC domain

van der Biezen EA, Jones JD, Curr Biol 1998;8:226-227.

30

365. Nucleoside diphosphate kinases active site

Nucleoside diphosphate kinases (EC 2.7.4.6) (NDK) [1] are enzymes required for the synthesis of nucleoside triphosphates (NTP) other than ATP. They provide NTPs for nucleic

acid synthesis, CTP for lipid synthesis, UTP for polysaccharide synthesis and GTP for protein elongation, signal transduction and microtubule polymerization. In eukaryotes, there seems to be a small family of NDK isozymes each of which acts in a different subcellular compartment and/or has a distinct biological function. Eukaryotic NDK isozymes are hexamers of two highly related chains (A and B) [2]. By random association (A6, A5B...AB5, B6), these two kinds of chain form isoenzymes differing in their isoelectric point. NDK are proteins of 17 Kd that act via a ping-pong mechanism in which a histidine residue is phosphorylated, by transfer of the terminal phosphate group from ATP. In the presence of magnesium, the phosphoenzyme can transfer its phosphate group to any NDP, to produce an NTP. NDK isozymes have been sequenced from prokaryotic and eukaryotic sources. It has also been shown [3] that the *Drosophila* awd (abnormal wing discs) protein, is a microtubule-associated NDK. Mammalian NDK is also known as metastasis inhibition factor nm23. The sequence of NDK has been highly conserved through evolution. There is a single histidine residue conserved in all known NDK isozymes, which is involved in the catalytic mechanism [2]. Our signature pattern contains this residue.

Consensus pattern: N-x(2)-H-[GA]-S-D-[SA]-[LIVMPKNE] [H is the putative active site residue]-

- [1] Parks R., Agarwal R. (In) The Enzymes (3rd edition) 8:307-334(1973).
- [2] Gilles A.-M., Presecan E., Vonica A., Lascu I. J. Biol. Chem. 266:8784-8789(1991).
- [3] Biggs J., Hersperger E., Steeg P.S., Liotta L.A., Shearn A. Cell 63:933-940(1990).

366. Nitrite and sulfite reductases iron-sulfur/siroheme-binding site (NIR_SIR)
Nitrite reductases (NiR) [1] catalyze the reduction of nitrite into ammonium, the second step in the assimilation of nitrate. There are two types of NiR: the higher plant chloroplastic form of NiR (EC 1.7.7.1) is a monomeric protein that uses reduced ferredoxin as the electron donor; while fungal and bacterial NiR (EC 1.6.6.4) are homodimeric proteins that uses NAD(P)H as the electron donor. Both forms of NiR contain a siroheme-Fe and iron-sulfur centers. Sulfite reductase (NADPH) (EC 1.8.1.2) (SIR) [2] is the bacterial enzyme that catalyzes the reduction of sulfite to sulfide. SIR is an oligomeric enzyme with a subunit composition of alpha(8)-beta(4), the alpha component is a flavoprotein (SIR-FP), while the

beta component is a siroheme, iron-sulfurprotein (SIR-HP). Sulfite reductase (ferredoxin) (EC 1.8.7.1) [3] is a cyanobacterial and plant monomeric enzyme that also catalyzes the reduction of sulfite to sulfide. Anaerobic sulfite reductase (EC 1.8.1.-) (ASR) [4], a bacterial enzyme that catalyzes the NADH-dependent reduction of sulfite to sulfide. ASR is an oligomeric enzyme composed of three different subunits. The C component (geneasrC) seems to be a siroheme, iron-sulfur protein. These enzymes share a region of sequence similarity in their C-terminal half; this region which spans about 80 amino acids includes four conserved cysteine residues. Two of the Cys are grouped together at the beginning of the domain, and the two others are grouped in the middle of the domain. The cysteines are involved in the binding of the iron-sulfur center; the last one also binds the siroheme group [2]. A signature pattern from the region around the second cluster of cysteines was derived.

Consensus pattern: [STV]-G-C-x(3)-C-x(6)-[DE]-[LIVMF]-[GAT]-[LIVMF] [The two C's are iron-sulfur ligands]-

[1] Campbell W.H., Kinghorn J.R. Trends Biochem. Sci. 15:315-319(1990).

[2] Crane B.R., Siegel L.M., Getzoff E.D. Science 270:59-67(1995).

[3] Gisselmann G., Klausmeier P., Schwenn J.D. Biochim. Biophys. Acta 1144:102-106(1993).

[4] Huang C.J., Barrett E.L. J. Bacteriol. 173:1544-1553(1991).

367. (NMT) Myristoyl-CoA:protein N-myristoyltransferase signatures. Myristoyl-CoA:protein N-myristoyltransferase (EC 2.3.1.97) (Nmt) [1] is the enzyme responsible for transferring a myristate group on the N-terminal glycine of a number of cellular eukaryotic and viral proteins. Nmt is a monomeric protein of about 50 to 60 Kd whose sequence appears to be well conserved. Two highly conserved regions have been developed as signature patterns. The first one is located in the central section, the second in the C-terminal part.

Consensus pattern: E-I-N-F-L-C-x-H-K-

Consensus pattern: K-F-G-x-G-D-G-

[1] Rudnick D.A., McWherter C.A., Gokel G.W., Gordon J.I. Adv. Enzymol. 67:375-430(1993).

5 368. ADP-glucose pyrophosphorylase signatures (NTP_transferase)

ADP-glucose pyrophosphorylase (glucose-1-phosphate adenylyltransferase) [1,2](EC 2.7.7.27) catalyzes a very important step in the biosynthesis of alpha 1,4-glucans (glycogen or starch) in bacteria and plants: synthesis of the activated glucosyl donor, ADP-glucose, from glucose-1-phosphate and ATP. ADP-glucose pyrophosphorylase is a tetrameric
10 allosterically regulated enzyme. It is a homotetramer in bacteria while in plant chloroplasts and amyloplasts, it is a heterotetramer of two different, yet evolutionary related, subunits. There are a number of conserved regions in the sequence of bacterial and plant ADP-glucose pyrophosphorylase subunits. Three of these regions were selected as signature patterns. The first two are N-terminal and have been proposed to be part of the allosteric and/or substrate-binding sites in the Escherichia coli enzyme (gene glgC). The third pattern corresponds to a
15 conserved region in the central part of the enzymes.

Consensus pattern: [AG]-G-G-x-G-[STK]-x-L-x(2)-L-[TA]-x(3)-A-x-P-A-[LV] -

Consensus pattern: W-[FY]-x-G-[ST]-A-[DNSH]-[AS]-[LIVMFYW]-

20 Consensus pattern: [APV]-[GS]-M-G-[LIVMN]-Y-[IVC]-[LIVMFY]-x(2)-[DENPHK] -

[1] Nakata P.A., Greene T.W., Anderson J.M., Smith-White B.J., Okita T.W., Preiss J. Plant Mol. Biol. 17:1089-1093(1991).

[2] Preiss J., Ball K., Hutney J., Smith-White B.J., Li. L., Okitsa T.W. Pure Appl. Chem.
25 63:535-544(1991).

369. Sodium/hydrogen exchanger family

Na/H antiporters are key transporters in maintaining the
30 pH of actively metabolizing cells. The molecular mechanisms of antiport are unclear.

These antiporters contain 10-12 transmembrane regions (M) at the amino-terminus and a large cytoplasmic region at the carboxyl

terminus. The transmembrane regions M3-M12 share identity with other members of the family. The M6 and M7 regions are highly conserved. Thus, this is thought to be the region that is involved in the transport of sodium and hydrogen ions. The cytoplasmic region has little similarity throughout the family.

[1] Dibrov P, Fliegel L; FEBS Lett 1998;424:1-5. [2] Orlowski J, Grinstein S; J Biol Chem 1997;272:22373-22376.[3] Numata M, Petrecca K, Lake N, Orlowski J; J Biol Chem 1998;273:6951-6959.

370. Sodium:sulfate symporter family signature (Na_sulph_syp)

Integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families currently consists of the following proteins: - Mammalian sodium/sulfate cotransporter [1]. - Mammalian renal sodium/dicarboxylate cotransporter [2], which transports succinate and citrate. - Mammalian intestinal sodium/dicarboxylate cotransporter. - Chlamydomonas reinhardtii putative sulfur deprivation response regulator SAC1 [3]. - Caenorhabditis elegans hypothetical proteins B0285.6, F31F6.6, K08E5.2 and R107.1. - Escherichia coli hypothetical protein yfbS. - Haemophilus influenzae hypothetical protein HI0608. - Synechocystis strain PCC 6803 hypothetical protein slI0640. - Methanococcus jannaschii hypothetical protein MJ0672. These transporters are proteins of from 430 to 620 amino acids which are highly hydrophobic and which probably contain about 12 transmembrane regions. As a signature pattern, a conserved region was selected which is located in or near the penultimate transmembrane region.

Consensus pattern: [STACP]-S-x(2)-F-x(2)-P-[LIVM]-[GSA]-x(3)-N-x-[LIVM]-V-

[1] Markovich D., Forgo J., Stange G., Biber J., Murer H. Proc. Natl. Acad. Sci. U.S.A. 90:8073-8077(1993).

[2] Pajor A.M. Am. J. Physiol. 270:642-648(1996).

[3] Davies J.P., Yildiz F.H., Grossman A. EMBO J. 15:2150-2159(1996).

371. NifU-like domain

This is an alignment of the carboxy-terminal domain. This is the only common region between the NifU protein from nitrogen-fixing bacteria and rhodobacterial species. The biochemical function of NifU is unknown [1].

Ouzounis C, Bork P, Sander C, Trends Biochem Sci 1994;19:199-200.

372. Nitrilases / cyanide hydratase signatures

Nitrilases (EC 3.5.5.1) are enzymes that convert nitriles into their corresponding acids and ammonia. They are widespread in microbes as well as in plants where they convert indole-3-acetonitrile to the hormone indole-3-acetic acid. A conserved cysteine has been shown [1,2] to be essential for enzyme activity; it seems to be involved in a nucleophilic attack on the nitrile carbon atom. Cyanide hydratase (EC 4.2.1.66) converts HCN to formamide. In phytopathogenic fungi, it is used to avoid the toxic effect of cyanide released by wounded plants [3]. The sequence of cyanide hydrolase is evolutionary related to that of nitrilases. Yeast hypothetical proteins YIL164c and YIL165c also belong to this family. As signature patterns for these enzymes, two conserved regions were selected. The first is located in the N-terminal section while the second, which contains the active site cysteine, is located in the central section.

Consensus pattern: G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P-

Consensus pattern: G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR] [C is the active site residue]-

[1] Kobayashi M., Izui H., Nagasawa T., Yamada H. Proc. Natl. Acad. Sci. U.S.A. 90:247-251(1993).

[2] Kobayashi M., Komeda H., Yanaka N., Nagasawa T., Yamada H. J. Biol. Chem. 267:20746-20751(1992).

[3] Wang P., Vanetten H.D. Biochem. Biophys. Res. Commun. 187:1048-1054(1992).

373. NusB family

The NusB protein is involved in the regulation of rRNA biosynthesis by transcriptional antitermination.

Huenges M, Rolz C, Gschwind R, Peteranderl R, Berglechner F, Richter G, Bacher A, Kessler H, Gemmecker G, EMBO J 1998;17:4092-4100.

5

374. (Neur Chan) Neurotransmitter-gated ion-channels signature

Neurotransmitter-gated ion-channels [1,2,3,4] provide the molecular basis for rapid signal transmission at chemical synapses. They are post-synaptic oligomeric transmembrane complexes that transiently form an ionic channel upon the binding of a specific neurotransmitter. Presently, the sequence of subunits from five types of neurotransmitter-gated receptors are known: - The nicotinic acetylcholine receptor (AChR), an excitatory cation channel. In the motor endplates of vertebrates, it is composed of four different subunits (alpha, beta, gamma and delta or epsilon) with a molar stoichiometry of 2:1:1:1. In neurones, the AChR receptor is composed of two different types of subunits: alpha and non-alpha (also called beta). Nicotinic AChRs are also found in invertebrates. - The glycine receptor, an inhibitory chloride ion channel. The glycine receptor is a pentamer composed of two different subunits (alpha and beta). - The gamma-aminobutyric-acid (GABA) receptor, which is also an inhibitory chloride ion channel. The quaternary structure of the GABA receptor is complex; at least four classes of subunits are known to exist (alpha, beta, gamma, and delta) and there are many variants in each class (for example: six variants of the alpha class have already been sequenced). - The serotonin 5HT3 receptor. Serotonin is a biogenic hormone that functions as a neurotransmitter, a hormone and a mitogen. There are seven major groups of serotonin receptors; six of these groups (5HT1, 5HT2, and 5HT4 to 5HT7) transduce extracellular signal by activating G proteins, while 5HT3 is a ligand-gated cation-specific ion channel which, when activated causes fast, depolarizing responses in neurons. - The glutamate receptor, an excitatory cation channel. Glutamate is the main excitatory neurotransmitter in the brain. At least three different types of glutamate receptors have been described and are named according to their selective agonists (kainate, N-methyl-D-aspartate (NMDA) and quisqualate). All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related. They are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions which form the ionic channel, followed by an intracellular region of variable length. A fourth

hydrophobic region is found at the C-terminal of the sequence. The sequence of subunits from the AchR, GABA, 5HT₃, and Gly receptors are clearly evolutionary related and share many regions of sequence similarities. These sequence similarities are either absent or very weak in the Glu receptors. In the N-terminal extracellular domain of AchR/GABA/5HT₃/Gly

5 receptors, there are two conserved cysteine residues, which, in AchR, have been shown to form a disulfide bond essential to the tertiary structure of the receptor. A number of amino acids between the two disulfide-bonded cysteines are also conserved. Therefore this region was used as a signature pattern for this subclass of proteins.

10 Consensus pattern: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C [The two C's are linked by a disulfide bond]-

[1] Stroud R.M., McCarthy M.P., Shuster M. *Biochemistry* 29:11009-11023(1990).

[2] Betz H. *Neuron* 5:383-392(1990).

15 [3] Dingleline R., Myers S.J., Nicholas R.A. *FASEB J.* 4:2632-2645(1990).

[4] Barnard E.A. *Trends Biochem. Sci.* 17:368-374(1992).

375. Orotidine 5'-phosphate decarboxylase active site

20 Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMPdecase) [1,2] catalyzes the last step in the de novo biosynthesis of pyrimidines, the decarboxylation of OMP into UMP. In higher eukaryotes OMPdecase is part, with orotatephosphoribosyltransferase, of a bifunctional enzyme, while the prokaryotic and fungal OMPdecases are monofunctional protein. Some parts of the sequence of OMPdecase are well conserved across species. The best conserved

25 region is located in the N-terminal half of OMPdecases and is centered around a lysine residue which is essential for the catalytic function of the enzyme. This region has been developed as a signature pattern.

30 Consensus pattern: [LIVMFTA]-[LIVMF]-x-D-x-K-x(2)-D-I-[GP]-x-T-[LIVMTA] [K is the active site residue]-

[1] Jacquet M., Guilbaud R., Garreau H. *Mol. Gen. Genet.* 211:441-445(1988).

[2] Kimsey H.H., Kaiser D. *J. Biol. Chem.* 267:819-824(1992).

376. ATP synthase delta (OSCP) subunit signature

ATP synthase (proton-translocating ATPase) (EC 3.6.1.34) [1,2] is a component of the cytoplasmic membrane of eubacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The ATPase complex is composed of an oligomeric transmembrane sector, called CF(0), which acts as a proton channel, and a catalytic core, termed coupling factor CF(1).

One of the subunits of the ATPase complex, known as subunit delta in bacteria and chloroplasts or the Oligomycin Sensitivity Conferral Protein (OSCP) in mitochondria, seems to be part of the stalk that links CF(0) to CF(1). It either transmits conformational changes from CF(0) into CF(1) or is involved in proton conduction [3].

The different delta/OSCP subunits are proteins of approximately 200 amino-acid residues - once the transit peptide has been removed in the chloroplast and mitochondrial forms - which show only moderate sequence homology.

The signature pattern used to detect ATPase delta/OSCP subunits is based on a conserved region in the C-terminal section of these proteins.

Consensus pattern: [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x-[GSA]-G-[LIVMFYGA]-x-[LIVM]-[KRHENQ]-x-[GSEN]

[1] Futai M., Noumi T., Maeda M. *Annu. Rev. Biochem.* 58:111-136(1989).

[2] Senior A.E. *Physiol. Rev.* 68:177-231(1988).

[3] Engelbrecht S., Junge W. *Biochim. Biophys. Acta* 1015:379-390(1990).

377. Aspartate and ornithine carbamoyltransferases signature

Aspartate carbamoyltransferase (EC 2.1.3.2) (ATCase) catalyzes the conversion of aspartate and carbamoyl phosphate to carbamoylaspartate, the second step in the de novo biosynthesis of pyrimidine nucleotides [1]. In prokaryotes ATCase consists of two subunits: a catalytic chain (gene *pyrB*) and a regulatory chain (gene *pyrI*), while in eukaryotes it is a domain in a multi-

functional enzyme (called URA2 in yeast, rudimentary in *Drosophila*, and CAD in mammals [2]) that also catalyzes other steps of the biosynthesis of pyrimidines.

Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase) catalyzes the conversion of ornithine and carbamoyl phosphate to citrulline. In mammals this enzyme participates in the urea cycle [3] and is located in the mitochondrial matrix. In prokaryotes and eukaryotic microorganisms it is involved in the biosynthesis of arginine. In some bacterial species it is also involved in the degradation of arginine [4] (the arginine deaminase pathway).

It has been shown [5] that these two enzymes are evolutionary related. The predicted secondary structure of both enzymes are similar and there are some regions of sequence similarities. One of these regions includes three residues which have been shown, by crystallographic studies [6], to be implicated in binding the phosphoryl group of carbamoyl phosphate.

This region was selected as a signature for these enzymes.

Consensus pattern: F-x-[EK]-x-S-[GT]-R-T[S, R, and the 2nd T bind carbamoyl phosphate]

-Note: the residue in position 3 of the pattern allows to distinguish between an ATCase (Glu) and an OTCase (Lys).

[1] Lerner C.G., Switzer R.L. J. Biol. Chem. 261:11156-11165(1986).

[2] Davidson J.N., Chen K.C., Jamison R.S., Musmanno L.A., Kern C.B. BioEssays 15:157-164(1993).

[3] Takiguchi M., Matsubasa T., Amaya Y., Mori M. BioEssays 10:163-166(1989).

[4] Baur H., Stalon V., Falmagne P., Luethi E., Haas D. Eur. J. Biochem. 166:111-117(1987).

[5] Houghton J.E., Bencini D.A., O'Donovan G.A., Wild J.R. Proc. Natl. Acad. Sci. U.S.A. 81:4864-4868(1981).

[6] Ke H.-M., Honzatko R.B., Lipscomb W.N. Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).

Oleosins [1] are the proteinaceous components of plants' lipid storage bodies called oil bodies. Oil bodies are small droplets (0.2 to 1.5 μm in diameter) containing mostly triacylglycerol that are surrounded by a phospholipid/oleosin annulus. Oleosins may have a structural role in stabilizing the lipid body during desiccation of the seed, by preventing coalescence of the oil. They may also provide recognition signals for specific lipase anchorage in lipolysis during seedling growth. Oleosins are found in the monolayer lipid/water interface of oil bodies and probably interact with both the lipid and phospholipid moieties.

Oleosins are proteins of 16 Kd to 24 Kd and are composed of three domains: an N-terminal hydrophilic region of variable length (from 30 to 60 residues); a central hydrophobic domain of about 70 residues and a C-terminal amphipathic region of variable length (from 60 to 100 residues). The central hydrophobic domain is proposed to be made up of beta-strand structure and to interact with the lipids [2]. It is the only domain whose sequence is conserved and therefore a section from that domain was selected as a signature pattern.

Consensus pattern: [AG]-[ST]-x(2)-[AG]-x(2)-[LIVM]-[SAD]-T-P-[LIVMF](4)-F-S-P-[LIVM](3)-P-A

- [1] Murphy D.J., Keen J.N., O'Sullivan J.N., Au D.M.Y., Edwards E.-W., Jackson P.J., Cummins I., Gibbons T., Shaw C.H., Ryan A.J. *Biochim. Biophys. Acta* 1088:86-94(1991).
- [2] Tzen J.T.C., Lie G.C., Huang A.H.C. *J. Biol. Chem.* 267:15626-15634(1992).

379. (Orbi VP5) Orbivirus outer capsid protein VP5

This paper shows the location of the different capsid proteins and their relation to each other.

- [1] Schoehn G, Moss SR, Nuttall PA, Hewat EA; *Virology* 1997;235:191-200.

380. Orn/DAP/Arg decarboxylases family 2 signatures

Pyridoxal-dependent decarboxylases acting on ornithine, lysine, arginine and related substrates can be classified into two different families on the basis of sequence similarities [1,2,3]. The second family consists of:

- Eukaryotic ornithine decarboxylase (EC 4.1.1.17) (ODC). ODC catalyzes the transformation of ornithine into putrescine.
- Prokaryotic diaminopimelic acid decarboxylase (EC 4.1.1.20) (DAPDC). DAPDC catalyzes the conversion of diaminopimelic acid into lysine; the last step in the biosynthesis of lysine.
- *Pseudomonas syringae* pv. *tabaci* protein *tabA*. *tabA* is probably involved in the biosynthesis of tabtoxin and is highly similar to DAPDC.
- Bacterial and plant biosynthetic arginine decarboxylase (EC 4.1.1.19) (ADC). ADC catalyzes the transformation of arginine into agmatine, the first step in the biosynthesis of putrescine from arginine.

The above proteins, while most probably evolutionary related, do not share extensive regions of sequence similarities. Two of the conserved regions were selected as signature patterns. The first pattern contains a conserved lysine residue which is known, in mouse ODC [4], to be the site of attachment of the pyridoxal-phosphate group. The second pattern contains a stretch of three consecutive glycine residues and has been proposed to be part of a substrate-binding region [5].

These enzymes are collectively known as group IV decarboxylases [3].

Consensus pattern: [FY]-[PA]-x-K-[SACV]-[NHCLFW]-x(4)-[LIVMF]-[LIVMTA]-x(2)-[LIVMA]-x(3)-[GTE] [K is the pyridoxal-P attachment site]

Consensus pattern: [GS]-x(2,6)-[LIVMSCP]-x(2)-[LIVMF]-[DNS]-[LIVMCA]-G-G-G-[LIVMFY]-[GSTPCEQ]

[1] Bairoch A. Unpublished observations (1993).

[2] Martin C., Cami B., Yeh P., Stragier P., Parsot C., Patte J.-C. Mol. Biol. Evol. 5:549-559(1988).

[3] Sandmeier E., Hale T.L., Christen P. Eur. J. Biochem. 221:997-1002(1994).

[4] Poulin R., Lu L., Ackermann B., Bey P., Pegg A.E. J. Biol. Chem. 267:150-158(1992).

[5] Moore R.C., Boyle S.M. J. Bacteriol. 172:4631-4640(1990).

381. Osteopontin signature

5 Osteopontin is an acidic phosphorylated glycoprotein of about 40 Kd which is abundant in the mineral matrix of bones and which binds tightly to hydroxyapatite [1,2,3]. It is suggested that osteopontin might function as a cell attachment factor and could play a key role in the adhesion of osteoclasts to the mineral matrix of bone.

10 Osteopontin-K is a kidney protein which is highly similar to osteopontin and probably also involved in cell-adhesion.

As a signature pattern a highly conserved region located at the N-terminal extremity of the mature protein was selected.

15 Consensus pattern: [KQ]-x-[TA]-x(2)-[GA]-S-S-E-E-K

[1] Butler W.T. Connect. Tissue Res. 23:123-36(1989).

[2] Gorski J.P. Calcif. Tissue Int. 50:391-396(1992).

[3] Denhardt D.T., Guo X. FASEB J. 7:1475-1482(1993).

20

382. Oxysterol-binding protein family signature

A number of eukaryotic proteins that seem to be involved with sterol synthesis and/or its regulation have been found [1] to be evolutionary related:

25 - Mammalian oxysterol-binding protein (OSBP). A protein of about 800 amino-acid residues that binds a variety of oxysterols: oxygenated derivatives of cholesterol. OSBP seems to play a complex role in the regulation of sterol metabolism.

30 - Yeast proteins HES1 and KES1; highly related proteins of 434 residues that seem to play a role in ergosterol synthesis.

- Yeast OSH1, a protein of 859 residues that also plays a role in ergosterol synthesis. - Yeast hypothetical protein YHR001w (437 residues).

- Yeast hypothetical protein YHR073w (996 residues).

360

- Yeast hypothetical protein YKR003w (448 residues).

All these proteins contain a moderately conserved domain of about 250 residues located in the C-terminal half of OBSP, OSH1 and YHR073w and in the central section of the other proteins. As a signature pattern, the best conserved part was selected of this domain, a region that contains a conserved pentapeptide.

Consensus pattern: E-[KQ]-x-S-H-[HR]-P-P-x-[STACF]-A

10 [1] Jiang B., Brown J.L., Sheraton J., Fortin N., Bussey H. Yeast 10:341-353(1994).

383. FMN oxidoreductase

384. Oxidoreductase FAD/NAD-binding domain

Number of members: 250

[1]

Medline: 92084635

The sequence of squash NADH:nitrate reductase and its relationship to the sequences of other flavoprotein oxidoreductases. A family of flavoprotein pyridine nucleotide cytochrome reductases.

Hyde GE, Crawford NM, Campbell W;

J Biol Chem 1991;266:23542-23547.

[2]Medline: 95111952

Crystal structure of the FAD-containing fragment of corn nitrate reductase at 2.5 Å resolution: relationship to other flavoprotein reductases.

Lu G, Campbell WH, Schneider G, Lindqvist Y;

Structure 1994;2:809-821.

385. (oxidored molyb) Eukaryotic molybdopterin oxidoreductases signature
A number of different eukaryotic oxidoreductases that require and bind a
molybdopterin cofactor have been shown [1] to share a few regions of sequence
similarity. These enzymes are:

5 - Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of
xanthine to uric acid with the concomitant reduction of NAD. Structurally,
this enzyme of about 1300 amino acids consists of at least three distinct
domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain
(see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mo-
pterin domain.

10 - Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into
acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its
sequence and domain structure.

15 - Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate
to nitrite. Structurally, this enzyme of about 900 amino acids consists of
an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding
domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome
reductase domain.

20 - Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to
sulfate. Structurally, this enzyme of about 460 amino acids consists of an
N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.

25 There are a few conserved regions in the sequence of the molybdopterin-binding
domain of these enzymes. The pattern used to detect these proteins is based
on one of them. It contains a cysteine residue which could be involved in
binding the molybdopterin cofactor.

Consensus pattern: [GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-
[DEN]-R-x(2)-[DE]

30 [1] Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle W.A., Bray
R.C. *Biochim. Biophys. Acta* 1057:157-185(1991).

386. (Oxidored q1) NADH-Ubiquinone/plastoquinone (complex I), various chains

This family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane. Number of members: 1824

[1]

Medline: 93110040

The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. Walker JE; Q Rev Biophys 1992;25:253-324.

387. (oxidored q3) NADH-ubiquinone/plastoquinone oxidoreductase chain 6. 179 members.

388. (oxidored q5) NADH-ubiquinone oxidoreductase chain 4, amino terminus

[1] Walker JE ; Q Rev Biophys 1992;25:253-324.

389. (oxidored q6) Respiratory-chain NADH dehydrogenase 20 Kd subunit signature Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase).

Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 20 Kd (in mammals) [3], which is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 20 Kd subunit has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in *Neurospora crassa*.
- Mitochondrial encoded in *Paramecium* (gene *psbG*).
- Chloroplast encoded in various higher plants (gene *ndhK* or *psbG*).

The 20 Kd subunit is highly similar to [4]:

- *Synechocystis* strain PCC 6803 proteins *psbG1* and *psbG2*.

- Subunit B of Escherichia coli NADH-ubiquinone oxidoreductase (gene nuoB).
- Subunit NQO6 of Paracoccus denitrificans NADH-ubiquinone oxidoreductase.
- Subunit 7 of Escherichia coli formate hydrogenlyase (gene hycG).
- Subunit I of Escherichia coli hydrogenase-4 (gene hylI).

5 As as signature pattern a highly conserved region was selected, located in the central section of this subunit and which contains a conserved cysteine that is probably involved in the binding of the 4Fe-4S center.

Consensus pattern: [GN]-x-D-[KRST]-[LIVMF](2)-P-[IV]-D-[LIVMFYW](2)-x-P-x-C-P-
10 [PT] [The C is a putative 4Fe-4S ligand]

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

[3] Arizmendi J.M., Runswick M.J., Skehel J.M., Walker J.E. FEBS Lett. 301:237-
242(1992).

15 [4] Weidner U., Geier S., Ptöck A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

390. p53 tumor antigen signature

20 The p53 tumor antigen [1 to 5, E1,E2] is a protein found in increased amounts in a wide variety of transformed cells. It is also detectable in many proliferating nontransformed cells, but it is undetectable or present at low levels in resting cells. It is frequently mutated or inactivated in many types of cancer. p53 seems to act as a tumor suppressor in some, but probably not
25 all, tumor types. p53 is probably involved in cell cycle regulation, and may be a trans-activator that acts to negatively regulate cellular division by controlling a set of genes required for this process.

p53 is a phosphoprotein of about 390 amino acids which can be subdivided into four domains: a highly charged acidic region of about 75 to 80 residues, a
30 hydrophobic proline-rich domain (position 80 to 150), a central region (from 150 to about 300), and a highly basic C-terminal region. The sequence of p53 is well conserved in vertebrate species; attempts to identify p53 in other eukaryotic phylum has so far been unsuccessful.

As a signature pattern for p53 a perfectly conserved stretch of 13 residues located in the central region of the protein was selected. This region, known as domain IV in [3], is involved (along with an adjacent region) in the binding of the large T antigen of SV40. In man this region is the focus of a variety of point mutations in cancerous tumors.

Consensus pattern: M-C-N-S-S-C-M-G-M-N-R-R

[1] Levine A.J., Momand J., Finlay C.A. Nature 351:453-456(1991).

[2] Levine A.J., Momand J. Biochim. Biophys. Acta 1032:119-136(1990).

[3] Soussi T., Caron De Fromentel C., May P. Oncogene 5:945-952(1990).

[4] Lane D.P., Benchimol S. Genes Dev. 4:1-8(1990).

[5] Ulrich S.J., Anderson C.W., Mercer W.E., Appella E. J. Biol. Chem. 267:15259-15262(1992).

391. (P5CR) Delta 1-pyrroline-5-carboxylate reductase signature

Delta 1-pyrroline-5-carboxylate reductase (P5CR) (EC 1.5.1.2) [1,2] is the enzyme that catalyzes the terminal step in the biosynthesis of proline from glutamate, the NAD(P) dependent oxidation of 1-pyrroline-5-carboxylate into proline.

The sequences of P5CR from eubacteria (gene proC), archaebacteria and eukaryotes show only a moderate level of overall similarity. As a signature pattern, the best conserved region located in the C-terminal section of P5CR was selected.

Consensus pattern: [PALF]-x(2,3)-[LIV]-x(3)-[LIVM]-[STAC]-[STV]-x-[GAN]-G-x-T-x(2)-[AG]-[LIV]-x(2)-[LMF]-[DENQK]

[1] Delauney A.J., Verma D.P. Mol. Gen. Genet. 221:299-305(1990).

[2] Savioz A., Jeenes D.J., Kocher H.P., Haas D. Gene 86:107-111(1990).

392. Poly-adenylate binding protein, unique domain.

393. (PAL) Phenylalanine and histidine ammonia-lyases active site

Phenylalanine ammonia-lyase (EC 4.3.1.5) (PAL) is a key enzyme of plant and
 5 fungi phenylpropanoid metabolism which is involved in the biosynthesis of a
 wide variety of secondary metabolites such as flavanoids, furanocoumarin
 phytoalexins and cell wall components. These compounds have many important
 roles in plants during normal growth and in responses to environmental stress.
 PAL catalyzes the removal of an ammonia group from phenylalanine to form
 10 trans-cinnamate.

Histidine ammonia-lyase (EC 4.3.1.3) (histidase) catalyzes the first step in
 histidine degradation, the removal of an ammonia group from histidine to
 produce urocanic acid.

The two types of enzymes are functionally and structurally related [1]. They
 15 are the only enzymes which are known to have the modified amino acid dehydro-
 alanine (DHA) in their active site. A serine residue has been shown [2,3,4] to
 be the precursor of this essential electrophilic moiety. The region around
 this active site residue is well conserved and can be used as a signature
 pattern.

20 Consensus pattern: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SA] [S is
 the active site residue]

- [1] Taylor R.G., Lambert M.A., Sexsmith E., Sadler S.J., Ray P.N., Mahuran D.J., McInnes
 25 R.R. J. Biol. Chem. 265:18192-18199(1990).
 [2] Langer M., Reck G., Reed J., Retey J. Biochemistry 33:6462-6467(1994).
 [3] Schuster B., Retey J. FEBS Lett. 349:252-254(1994).
 [4] Taylor R.G., McInnes R.R. J. Biol. Chem. 269:27473-27477(1994).

394. PAS domain

30 -!- CAUTION. This family does not currently match all known
 examples of PAS domains.

PAS motifs appear in archaea, eubacteria and eukarya. Probably the most surprising identification of a PAS domain was that in EAG-like K⁺-channels[1,3].

Number of members: 308

[1]

Medline: 97446881

PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox.

Zhulin IB, Taylor BL, Dixon R;

Trends Biochem Sci 1997;22:331-333.

[2]Medline: 95275818

1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore.

Borgstahl GE, Williams DR, Getzoff ED;

Biochemistry 1995;34:6278-6287.

[3]Medline: 98044337

PAS: a multifunctional domain family comes to light.

Ponting CP, Aravind L;

Curr Biol 1997;7:674-677.

395. (PBP) Phosphatidylethanolamine-binding protein family signature

Mammalian phosphatidylethanolamine-binding protein (also known as basic cytosolic 21 Kd protein) is a 186 residue protein found in a variety of tissues [1]. It binds hydrophobic ligands, such as phosphatidylethanolamine, but also seems [2] to bind nucleotides such as GTP and FMN, it is suggested that it could act in membrane remodeling during growth and maturation. This protein belongs to a family that also includes:

- Drosophila antennal protein A5, a putative odorant-binding protein.
- Onchocerca volvulus antigen Ov-16 and the related proteins D1, D2 and D3.
- Plasmodium falciparum putative phosphatidylethanolamine-binding protein.
- Toxocara canis secreted antigen TES-26. This larval protein has been shown to bind phosphatidylethanolamine.

367

- Yeast protein DKA1 (also known as NSP1 or TFS1). The function of this protein is not very clear.
- Yeast hypothetical protein YLR179C.
- *Caenorhabditis elegans* hypothetical protein F40A3.3.

As a signature pattern, the best conserved region was selected which is located in the end of the first third of the sequence of these proteins.

Consensus pattern: [FYL]-x-[LV]-[LIVF]-x-[TIV]-[DC]-P-D-x-P-[SN]-x(10)-H

- [1] Seddiqi N., Bollengier F., Alliel P.M., Perin J.P., Bonnet F., Bucquoy S., Jolles P., Schoentgen F. J. Mol. Evol. 39:655-660(1994).
- [2] Schoentgen F., Jolles P. FEBS Lett. 369:22-6(1995).

396. PCI domain

This domain has also been called the PINT motif (Proteasome, Int-6, Nip-1 and TRIP-15) [1].

Number of members: 49

[1]

Medline: 98308842

The PCI domain: a common theme in three multiprotein complexes.

Hofmann K, Bucher P;

Trends Biochem Sci 1998;23:204-205.

[2]Medline: 98266368

Homologues of 26S proteasome subunits are regulators of transcription and translation.

Aravind L, Ponting CP;

Protein Sci 1998;7:1250-1254.

397. (PCMT) Protein-L-isoaspartate (D-aspartate) O-methyltransferase signature. Protein-L-isoaspartate (D-aspartate) O-methyltransferase (EC 2.1.1.77) (PCMT)[1] (which is also known as L-isoaspartyl protein carboxyl methyltransferase) is an enzyme that catalyzes the

transfer of a methyl group from S-adenosylmethionine to the free carboxyl groups of D-aspartyl or L-aspartyl residues in a variety of peptides and proteins. The enzyme does not act on normal L-aspartyl residues L-isoaspartyl and D-aspartyl are the products of the spontaneous de amidation and/or isomerization of normal L-aspartyl and L-asparaginyl residues in proteins. PCMT plays a role in the repair and/or degradation of these damaged proteins; the enzymatic methyl esterification of the abnormal residues can lead to their conversion to normal L-aspartyl residues. PCMT is a well-conserved and widely distributed cytosolic protein of about 24Kd. As a signature pattern, a conserved region in the central part of this enzyme has been developed.

Consensus pattern: [GSA]-D-G-x(2)-G-[FYWV]-x(3)-[AS]-P-[FY]-[DN]-x-I -

[1] Kagan R.M., McFadden H.J., McFadden P.N., O'Connor C., Clarke S. Comp. Biochem. Physiol. 117b:379-385(1997).

398. (PCNA) Proliferating cell nuclear antigen signatures

Proliferating cell nuclear antigen (PCNA) [1,2] is a protein involved in DNA replication by acting as a cofactor for DNA polymerase delta, the polymerase responsible for leading strand DNA replication.

A similar protein exists in yeast (gene POL30) [3] and is associated with polymerase III, the yeast analog of polymerase delta. In baculoviruses the ETL protein has been shown [4] to be highly related to PCNA and is probably associated with the viral encoded DNA polymerase. An homolog of PCNA is also found in archebacteria.

As signatures for this family of proteins, two conserved regions were selected located in the N-terminal section. The second one has been proposed to bind DNA.

Consensus pattern: [GA]-[LIVMF]-x-[LIVMA]-x-[SAV]-[LIVM]-D-x-[NSAE]-[HKR]-[VI]-x-[LY]-[VGA]-x-[LIVM]-x-[LIVM]-x(4)-F

-Consensus pattern: [RKA]-C-[DE]-[RH]-x(3)-[LIVMF]-x(3)-[LIVM]-x-[SGAN]-[LIVMF]-x-K-[LIVMF](2)

- [1] Bravo R., Frank R., Blundell P.A., McDonald-Bravo H. *Nature* 326:515-517(1987).
 [2] Suzuka I., Hata S., Matsuoka M., Kosugi S., Hashimoto J. *Eur. J. Biochem.* 195:571-575(1991). [3] Bauer G.A., Burgess P.M.J. *Nucleic Acids Res.* 18:261-265(1990).
 5 [4] O'Reilly D.R., Crawford A.M., Miller L.K. *Nature* 337:606-606(1989).

399. (PDT) Prephenate dehydratase signatures

Prephenate dehydratase (EC 4.2.1.51) (PDT) catalyzes the decarboxylation of
 10 prephenate into phenylpyruvate. In microorganisms PDT is involved in the
 terminal pathway of the biosynthesis of phenylalanine. In some bacteria such
 as *Escherichia coli* PDT is part of a bifunctional enzyme (P-protein) that also
 catalyzes the transformation of chorismate into prephenate (chorismate
 mutase) while in other bacteria it is a monofunctional enzyme. The sequence of
 15 monofunctional PDT align well with the C-terminal part of that of P-proteins
 [1].

As signature patterns for PDT two conserved regions were selected. The first
 region contains a conserved threonine which has been said to be essential for
 the activity of the enzyme in *E. coli*. The second region includes a conserved
 20 glutamate. Both regions are in the C-terminal part of PDT.

Consensus pattern: [FY]-x-[LIVM]-x(2)-[LIVM]-x(5)-[DN]-x(5)-T-R-F-[LIVMW]-x-
 [LIVM]

- 25 [1] Fischer R.S., Zhao G., Jensen R.A. *J. Gen. Microbiol.* 137:1293-1301(1991).

400. PDZ domain (Also known as DHR or GLGF).

PDZ domains are found in diverse signaling proteins.

- 30 [1] Ponting CP, Phillips C, Davies KE, Blake DJ

Bioessays 1997;19:469-479. [2] Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R;
Cell. 1996;85:1067-1076. [3] Ponting CP; *Protein Sci* 1997;6:464-468.

401. (PPDK_N_term) PEP-utilizing enzymes signatures

A number of enzymes that catalyze the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) via a phospho-histidine intermediate have been shown to be structurally related [1,2,3,4]. These enzymes are:

- Pyruvate,orthophosphate dikinase (EC 2.7.9.1) (PPDK). PPDK catalyzes the reversible phosphorylation of pyruvate and phosphate by ATP to PEP and diphosphate. In plants PPDK function in the direction of the formation of PEP, which is the primary acceptor of carbon dioxide in C4 and crassulacean acid metabolism plants. In some bacteria, such as *Bacteroides symbiosus*, PPDK functions in the direction of ATP synthesis.

- Phosphoenolpyruvate synthase (EC 2.7.9.2) (pyruvate,water dikinase). This enzyme catalyzes the reversible phosphorylation of pyruvate by ATP to form PEP, AMP and phosphate, an essential step in gluconeogenesis when pyruvate and lactate are used as a carbon source.

- Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9). This is the first enzyme of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS), a major carbohydrate transport system in bacteria. The PTS catalyzes the phosphorylation of incoming sugar substrates concomitant with their translocation across the cell membrane. The general mechanism of the PTS is the following: a phosphoryl group from PEP is transferred to enzyme-I (EI) of PTS which in turn transfers it to a phosphoryl carrier protein (HPr). Phospho-HPr then transfers the phosphoryl group to a sugar-specific permease.

All these enzymes share the same catalytic mechanism: they bind PEP and transfer the phosphoryl group from it to a histidine residue. The sequence around that residue is highly conserved and can be used as a signature pattern for these enzymes. As a second signature pattern a conserved region was selected in the C-terminal part of the PEP-utilizing enzymes. The biological significance of this region is not yet known.

Consensus pattern: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG] [H is phosphorylated]

-Consensus pattern: [DEQSK]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-
[LIVMFYGT]-[STALIV]-[LIVMF]-[GAS]-x(2)-R

[1] Reizer J., Hoischen C., Reizer A., Pham T.N., Saier M.H. Jr. Protein Sci. 2:506-521(1993).

[2] Reizer J., Reizer A., Merrick M.J., Plunkett G. III, Rose D.J., Saier M.H. Jr. Gene 181:103-108(1996).

[3] Pocalyko D.J., Carroll L.J., Martin B.M., Babbitt P.C., Dunaway-Mariano D. Biochemistry 29:10757-10765(1990).

[4] Niersbach M., Kreuzaler F., Geerse R.H., Postma P., Hirsch H.J. Mol. Gen. Genet. 232:332-336(1992).

402. (PEPCK ATP) Phosphoenolpyruvate carboxykinase (ATP) signature
Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) (PEPCK) [1] catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate while hydrolyzing ATP, a rate limiting step in gluconeogenesis (the biosynthesis of glucose).

The sequence of this enzyme has been obtained from Escherichia coli, yeast, and Trypanosoma brucei; these three sequences are evolutionary related and share many regions of similarity. As a signature pattern a highly conserved region was selected that contains four acidic residues and which is located in the central part of the enzyme. The beginning of the pattern is located about 10 residues to the C-terminus of an ATP-binding motif 'A' (P-loop) (see <PDOC00017>) and is also part of the ATP-binding domain [2].

Consensus pattern: L-I-G-D-D-E-H-x-W-x-[DE]-x-G-[IV]-x-N

-Note: phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) an enzyme that catalyzes the same reaction, but using GTP instead of ATP, is not related to the above enzyme (see <PDOC00421>).

[1] Medina V., Pontarollo R., Glaeske D., Tabel H., Goldie H. J. Bacteriol. 172:7151-7156(1990).

[2] Matte A., Goldie H., Sweet R.M., Delbaere L.T.J. J. Mol. Biol. 256:126-143(1996).

403. (Pepcase) Phosphoenolpyruvate carboxylase active sites. Phosphoenolpyruvate
 5 carboxylase (EC 4.1.1.31) (PEPcase) catalyzes the irreversible beta-carboxylation of
 phosphoenolpyruvate by bicarbonate to yield oxaloacetate and phosphate. The enzyme is
 found in all plants and in a variety of microorganisms. A histidine [1] and a lysine [2] have
 been implicated in the catalytic mechanism of this enzyme; the regions around these active
 site residues are highly conserved in PEPcase from various plants, bacteria and cyanobacteria
 10 and can be used as a signature patterns for this type of enzyme.

Consensus pattern: [VT]-x-T-A-H-P-T-[EQ]-x(2)-R-[KRH] [H is an active site residue]-
 Consensus pattern: [IV]-M-[LIVM]-G-Y-S-D-S-x-K-D-[STAG]-G [K is an active site
 residue]-

15 [1] Terada K., Izui K. Eur. J. Biochem. 202:797-803(1991).[2] Jiao J.-A., Podesta F.E.,
 Chollet R., O'Leary M.H., Andreo C.S. Biochim. Biophys. Acta 1041:291-295(1990).

20 404. PET112 family signature

The following proteins from eukaryotes, prokaryotes and archaeobacteria belong
 to the same family:

- Yeast mitochondrial protein PET112 [1], which plays an unknown role in the
 expression of mitochondrial genes, probably at the level of translation.
- 25 - *Aspergillus nidulans* mitochondrial protein nempA.
- *Bacillus subtilis* hypothetical protein yzdD.
- *Moraxella catarrhalis* hypothetical protein in bloR-1 3'region.
- *Mycoplasma genitalium* hypothetical protein MG100.
- *Methanococcus jannaschii* hypothetical proteins MJ0019 and MJ0160.

30 The size of these proteins range from 419 to 630 amino acids. As a signature
 pattern, a conserved region located in the N-terminal section was selected.

Consensus pattern: [DN]-x-[DN]-R-x(3)-P-L-[LIV]-E-[LIV]-x-[ST]-x-P

[1] Mulero J.J., Rosenthal J.K., Fox T.D. Curr. Genet. 25:299-304(1994).

405. (PFK) Phosphofructokinase signature

Phosphofructokinase (EC 2.7.1.11) (PFK) [1,2] is a key regulatory enzyme in the glycolytic pathway. It catalyzes the phosphorylation by ATP of fructose 6-phosphate to fructose 1,6-bisphosphate. In bacteria PFK is a tetramer of identical 36 Kd subunits. In mammals it is a tetramer of 80 Kd subunits. Each 80 Kd subunit consist of two homologous domains which are highly related to the bacterial 36 Kd subunits. In Human there are three, tissue-specific, types of PFK isozymes: PFKM (muscle), PFKL (liver), and PFKP (platelet). In yeast PFK is an octamer composed of four 100 Kd alpha chains (gene PFK1) and four 100 Kd beta chains (gene PFK2); like the mammalian 80 Kd subunits, the yeast 100 Kd subunits are composed of two homologous domains. As a signature pattern for PFK a region that contains three basic residues involved in fructose-6-phosphate binding was selected.

Consensus pattern: [RK]-x(4)-G-H-x-Q-[QR]-G-G-x(5)-D-R [The R/K, the H and the Q/R are involved in fructose-6-P binding]

-Note: Escherichia coli has two phosphofructokinase isozymes which are encoded by genes pfkA (major) and pfkB (minor). The pfkB isozyme is not evolutionary related to other prokaryotic or eukaryotic PFK's (see <PDOC00504>).

[1] Poorman R.A., Randolph A., Kemp R.G., Heinrikson R.L. Nature 309:467-469(1984).
[2] Heinisch J., Ritzel R.G., von Borstel R.C., Aguilera A., Rodicio R., Zimmermann F.K. Gene 78:309-321(1989).

406. (PGAM) Phosphoglycerate mutase family phosphohistidine signature

Phosphoglycerate mutase (EC 5.4.2.1) (PGAM) and bisphosphoglycerate mutase (EC 5.4.2.4) (BPGM) are structurally related enzymes which catalyze reactions involving the transfer of phospho groups between the three carbon atoms of

phosphoglycerate [1,2]. Both enzymes can catalyze three different reactions, although in different proportions:

- The isomerization of 2-phosphoglycerate (2-PGA) to 3-phosphoglycerate (3-PGA) with 2,3-diphosphoglycerate (2,3-DPG) as the primer of the reaction.
- 5 - The synthesis of 2,3-DPG from 1,3-DPG with 3-PGA as a primer.
- The degradation of 2,3-DPG to 3-PGA (phosphatase EC 3.1.3.13 activity).

In mammals, PGAM is a dimeric protein. There are two isoforms of PGAM: the M (muscle) and B (brain) forms. In yeast, PGAM is a tetrameric protein. BPGM is a dimeric protein and is found mainly in erythrocytes where it plays a major
10 role in regulating hemoglobin oxygen affinity as a consequence of controlling 2,3-DPG concentration.

The catalytic mechanism of both PGAM and BPGM involves the formation of a phosphohistidine intermediate [3].

The bifunctional enzyme 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase
15 (EC 2.7.1.105 and EC 3.1.3.46) (PF2K) [4] catalyzes both the synthesis and the degradation of fructose-2,6-bisphosphate. PF2K is an important enzyme in the regulation of hepatic carbohydrate metabolism. Like PGAM/BPGM, the fructose-2,6-bisphosphatase reaction involves a phosphohistidine intermediate and the phosphatase domain of PF2K is structurally related to PGAM/BPGM.

20 The bacterial enzyme alpha-ribazole-5'-phosphate phosphatase (gene cobC) which is involved in cobalamin biosynthesis also belongs to this family [5].

A signature pattern was built around the phosphohistidine residue.

Consensus pattern: [LIVM]-x-R-H-G-[EQ]-x(3)-N [H is the phosphohistidine residue]

25 -Note: some organisms harbor a form of PGAM independent of 2,3-DPG, this enzyme is not related to the family described above [6].

[1] Le Boulch P., Joulin V., Garel M.-C., Rosa J., Cohen-Solal M. Biochem. Biophys. Res. Commun. 156:874-881(1988).

30 [2] White M.F., Fothergill-Gilmore L.A. FEBS Lett. 229:383-387(1988).

[3] Rose Z.B. Meth. Enzymol. 87:43-51(1982).

[4] Bazan J.F., Fletterick R.J., Pilgis S.J. Proc. Natl. Acad. Sci. U.S.A. 86:9642-9646(1989).

[5] O'Toole G.A., Trzebiatowski J.R., Escalante-Semerena J.C. J. Biol. Chem. 269:26503-26511(1994).

[6] Grana X., De Lecea L., El-Maghrabi M.R., Urena J.M., Caellas C., Carreras J., Puigdomenech P., Pilakis S.J., Climent F. J. Biol. Chem. 267:12797-12803(1992).

5

407. (PGI) Phosphoglucose isomerase signatures

Phosphoglucose isomerase (EC 5.3.1.9) (PGI) [1,2] is a dimeric enzyme that catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate. PGI is involved in different pathways: in most higher organisms it is involved in glycolysis; in mammals it is involved in gluconeogenesis; in plants in carbohydrate biosynthesis; in some bacteria it provides a gateway for fructose into the Entner-Doudouff pathway. PGI has been shown [3] to be identical to neuroleukin, a neurotrophic factor which supports the survival of various types of neurons.

The sequence of PGI from many species ranging from bacteria to mammals is available and has been shown to be highly conserved. As signature patterns for this enzyme two conserved regions were selected, the first region is located in the central section of PGI, while the second one is located in its C-terminal section.

Consensus pattern: [DENS]-x-[LIVM]-G-G-R-[FY]-S-[LIVMT]-x-[STA]-[PSAC]-[LIVMA]-G

-Consensus pattern: [GS]-x-[LIVM]-[LIVMFYW]-x(4)-[FY]-[DN]-Q-x-G-V-E-x(2)-K

[1] Achari A., Marshall S.E., Muirhewad H., Palmieri R.H., Noltmann E.A. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 293:145-157(1981).

[2] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:544-545(1992).

[3] Faik P., Walker J.I.H., Redmill A.A.M., Morgan M.J. Nature 332:455-456(1988).

408. (PGK) Phosphoglycerate kinase signature

Phosphoglycerate kinase (EC 2.7.2.3) (PGK) [1] catalyzes the second step in

the second phase of glycolysis, the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate with generation of one molecule of ATP. PGK is found in all living organisms and its sequence has been highly conserved throughout evolution. It is a two-domain protein; each domain is composed of six repeats of an alpha/beta structural motif. As a signature pattern for PGK's, a conserved region in the N-terminal region was selected.

Consensus pattern: [KRHGTCVN]-[VT]-[LIVMF]-[LIVMC]-R-x-D-x-N-[SACV]-P

[1] Watson H.C., Littlechild J.A. Biochem. Soc. Trans. 18:187-190(1990).

409. (PGM PMM) Phosphoglucomutase and phosphomannomutase phosphoserine signature

- Phosphoglucomutase (EC 5.4.2.2) (PGM). PGM is an enzyme responsible for the conversion of D-glucose 1-phosphate into D-glucose 6-phosphate. PGM participates in both the breakdown and synthesis of glucose [1].
- Phosphomannomutase (EC 5.4.2.8) (PMM). PMM is an enzyme responsible for the conversion of D-mannose 1-phosphate into D-mannose 6-phosphate. PMM is required for different biosynthetic pathways in bacteria. For example, in enterobacteria such as *Escherichia coli* there are two different genes coding for this enzyme: *rfbK* which is involved in the synthesis of the O antigen of lipopolysaccharide and *cpsG* which is required for the synthesis of the M antigen capsular polysaccharide [2]. In *Pseudomonas aeruginosa* PMM (gene *algC*) is involved in the biosynthesis of the alginate layer [3] and in *Xanthomonas campestris* (gene *xanA*) it is involved in the biosynthesis of xanthan [4]. In *Rhizobium* strain *ngr234* (gene *noeK*) it is involved in the biosynthesis of the nod factor.
- Phosphoacetylglucosamine mutase (EC 5.4.2.3) which converts N-acetyl-D-glucosamine 1-phosphate into the 6-phosphate isomer.

The catalytic mechanism of both PGM and PMM involves the formation of a phosphoserine intermediate [1]. The sequence around the serine residue is well conserved and can be used as a signature pattern.

In addition to PGM and PMM there are at least three uncharacterized proteins that belong to this family [5,6]:

- Urease operon protein ureC from *Helicobacter pylori*.
- *Escherichia coli* protein mrsA.
- *Paramecium tetraurelia* parafusin, a phosphoglycoprotein involved in exocytosis.
- 5 - A *Methanococcus vannielii* hypothetical protein in the 3'region of the gene for ribosomal protein S10.

Consensus pattern: [GSA]-[LIVM]-x-[LIVM]-[ST]-[PGA]-S-H-x-P-x(4)-[GNHE] [S is the phosphoserine residue]

- 10 -Note: PMM from fungi do not belong to this family.

- [1] Dai J.B., Liu Y., Ray W.J. Jr., Konno M. J. Biol. Chem. 267:6322-6337(1992).
- [2] Stevenson G., Lee S.J., Romana L.K., Reeves P.R. Mol. Gen. Genet. 227:173-180(1991).
- 15 [3] Zielinski N.A., Chakrabarty A.M., Berry A. J. Biol. Chem. 266:9754-9763(1991).
- [4] Koeplin R., Arnold W., Hoette B., Simon R., Wang G., Puehler A. J. Bacteriol. 174:191-199(1992).
- [5] Bairoch A. Unpublished observations (1993).
- [6] Subramanian S.V., Wyroba E., Andersen A.P., Satir B.H. Proc. Natl. Acad. Sci. U.S.A.
- 20 91:9832-9836(1994).

410. PH domain profile

The 'pleckstrin homology' (PH) domain is a domain of about 100 residues that occurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton [1 to 7].

The function of this domain is not clear, several putative functions have been suggested: - binding to the beta/gamma subunit of heterotrimeric G proteins,

- binding to lipids, e.g. phosphatidylinositol-4,5-bisphosphate,

- binding to phosphorylated Ser/Thr residues,

- attachment to membranes by an unknown mechanism.

It is possible that different PH domains have totally different ligand requirements.

The 3D structure of several PH domains has been determined [8]. All known cases have a common structure consisting of two perpendicular anti-parallel beta sheets, followed by a C-terminal amphipathic helix. The loops connecting the beta-strands differ greatly in length, making the PH domain relatively difficult to detect. There are no totally invariant residues within the PH domain.

Proteins reported to contain one more PH domains belong to the following families:

- Pleckstrin, the protein where this domain was first detected, is the major substrate of protein kinase C in platelets. Pleckstrin is one of the rare proteins to contains two PH domains.
- Ser/Thr protein kinases such as the Act/Rac family, the beta-adrenergic receptor kinases, the mu isoform of PKC and the trypanosomal NrK A family.
- Tyrosine protein kinases belonging to the Btk/Itk/Tec subfamily.
- Insulin Receptor Substrate 1 (IRS-1).
- Regulators of small G-proteins like guanine nucleotide releasing factor GNRP (Ras-GRF) (which contains 2 PH domains), guanine nucleotide exchange proteins like vav, db1, SoS and yeast CDC24, GTPase activating proteins like rasGAP and BEM2/IPL2, and the human break point cluster protein bcr.
- Cytoskeletal proteins such as dynamin (see <PDOC00362>), Caenorhabditis elegans kinesin-like protein unc-104 (see <PDOC00343>), spectrin beta-chain, syntrophin (2 PH domains) and yeast nuclear migration protein NUM1.
- Mammalian phosphatidylinositol-specific phospholipase C (PI-PLC) (see <PDOC50007>) isoforms gamma and delta. Isoform gamma contains two PH domains, the second one is split into two parts separated by about 400 residues.
- Oxysterol binding proteins OSBP, yeast OSH1 and YHR073w.
- Mouse protein citron, a putative rho/rac effector that binds to the GTP-bound forms of rho and rac,
- Several yeast proteins involved in cell cycle regulation and bud formation like BEM2, BEM3, BUD4 and the BEM1-binding proteins BOI2 (BEB1) and BOI1 (BOB1).
- Caenorhabditis elegans protein MIG-10.
- Caenorhabditis elegans hypothetical proteins C04D8.1, K06H7.4 and ZK632.12.
- Yeast hypothetical proteins YBR129c and YHR155w.

The profile for the PH domain, which has been developed by Toby Gibson at the EMBL, covers the total length of domain. Several proteins contain large insertions in the PH domain and are thus difficult to detect with this profile. In some of these cases, the profile will align only to one half of the PH domain.

-Sequences known to belong to this class detected by the pattern: ALL. But it should be noted that while all sequences containing PH domains are detected, not all PH domains are. Some of the split domains lie below the cutoff threshold.

- [1] Mayer B.J., Ren R., Clark K.L., Baltimore D. Cell 73:629-630(1993).
- [2] Haslam R.J., Koide H.B., Hemmings B.A. Nature 363:309-310(1993).
- [3] Musacchio A., Gibson T.J., Rice P., Thompson J., Saraste M. Trends Biochem. Sci. 18:343-348(1993).
- [4] Gibson T.J., Hyvonen M., Musacchio A., Saraste M., Birney E. Trends Biochem. Sci. 19:349-353(1994).
- [5] Pawson T. Nature 373:573-580(1995).
- [6] Ingle E., Hemmings B.A. J. Cell. Biochem. 56:436-443(1994).
- [7] Saraste M., Hyvonen M. Curr. Opin. Struct. Biol. 5:403-408(1995).
- [8] Riddihough G. Nat. Struct. Biol. 1:755-757(1994).

411. PHD-finger

[1]

Medline: 95216093

The PHD finger: implications for chromatin-mediated transcriptional regulation.

Aasland R, Gibson TJ, Stewart AF;
Trends Biochem Sci 1995;20:56-59.

Number of members: 181

412. (PI-PLC-X) Phosphatidylinositol-specific phospholipase C profiles
Phosphatidylinositol-specific phospholipase C (EC 3.1.4.11), an eukaryotic

intracellular enzyme, plays an important role in signal transduction processes [1]. It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5-triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins [2 to 4].

In mammals, there are at least 6 different isoforms of PI-PLC, they differ in their domain structure, their regulation, and their tissue distribution. Lower eukaryotes also possess multiple isoforms of PI-PLC.

All eukaryotic PI-PLCs contain two regions of homology, sometimes referred to as 'X-box' and 'Y-box'. The order of these two regions is always the same (NH₂-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain (see <PDOC00380>) possibly involved in Ca-dependent membrane attachment.

Profile analysis shows that sequences with significant similarity to the X-box domain occur also in prokaryotic and trypanosome PI-specific phospholipases C. Apart from this region, the prokaryotic enzymes show no similarity to their eukaryotic counterparts.

Two profiles were developed, one covering the X-box, the other the Y-box.

[1] Meldrum E., Parker P.J., Carozzi A.

Biochim. Biophys. Acta 1092:49-71(1991).[2] Rhee S.G., Choi K.D.

Adv. Second Messenger Phosphoprotein Res. 26:35-61(1992).

[3] Rhee S.G., Choi K.D. J. Biol. Chem. 267:12393-12396(1992).

[4] Sternweis P.C., Smrcka A.V. Trends Biochem. Sci. 17:502-506(1992).

413. (PI-PLC-Y) Phosphatidylinositol-specific phospholipase C profiles
Phosphatidylinositol-specific phospholipase C (EC 3.1.4.11), an eukaryotic intracellular enzyme, plays an important role in signal transduction processes [1]. It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5-

triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins [2 to 4].

In mammals, there are at least 6 different isoforms of PI-PLC, they differ in their domain structure, their regulation, and their tissue distribution. Lower eukaryotes also possess multiple isoforms of PI-PLC.

All eukaryotic PI-PLCs contain two regions of homology, sometimes referred to as 'X-box' and 'Y-box'. The order of these two regions is always the same (NH₂-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain (see <PDOC00380>) possibly involved in Ca-dependent membrane attachment.

Profile analysis shows that sequences with significant similarity to the X-box domain occur also in prokaryotic and trypanosome PI-specific phospholipases C. Apart from this region, the prokaryotic enzymes show no similarity to their eukaryotic counterparts.

Two profiles were developed, one covering the X-box, the other the Y-box.

[1] Meldrum E., Parker P.J., Carozzi A.

Biochim. Biophys. Acta 1092:49-71(1991).[2] Rhee S.G., Choi K.D.

Adv. Second Messenger Phosphoprotein Res. 26:35-61(1992).

[3] Rhee S.G., Choi K.D. J. Biol. Chem. 267:12393-12396(1992).

[4] Sternweis P.C., Smrcka A.V. Trends Biochem. Sci. 17:502-506(1992).

414. (PK) Pyruvate kinase active site signature

Pyruvate kinase (EC 2.7.1.40) (PK) [1] catalyzes the final step in glycolysis, the conversion of phosphoenolpyruvate to pyruvate with the concomitant phosphorylation of ADP to ATP. PK requires both magnesium and potassium ions for its activity. PK is found in all living organisms. In vertebrates there are four, tissues specific, isozymes: L (liver), R (red cells), M1 (muscle,

heart, and brain), and M2 (early fetal tissues). In *Escherichia coli* there are two isozymes: PK-I (gene *pykF*) and PK-II (gene *pykA*). All PK isozymes seem to be tetramers of identical subunits of about 500 amino acid residues.

As a signature pattern for PK a conserved region was selected that includes a

lysine residue which seems to be the acid/base catalyst responsible for the interconversion of pyruvate and enolpyruvate, and a glutamic acid residue implicated in the binding of the magnesium ion.

Consensus pattern: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQHS]-
[GSTA]-[LIVM] [K is the active site residue] [E is a magnesium ligand]

[1] Muirhead H. *Biochem. Soc. Trans.* 18:193-196(1990).

415. (PLDc) Phospholipase D. Active site motif

Phosphatidylcholine-hydrolyzing phospholipase D (PLD) isoforms are activated by ADP-ribosylation factors (ARFs). PLD produces phosphatidic acid from phosphatidylcholine, which may be essential for the formation of certain types of transport vesicles or may be constitutive vesicular transport to signal transduction pathways.

PC-hydrolyzing PLD is a homologue of cardiolipin synthase, phosphatidylserine synthase, bacterial PLDs, and viral proteins.

Each of these appears to possess a domain duplication which is apparent by the presence of two motifs containing well-conserved histidine, lysine, and/or asparagine residues which may contribute to the active site.

aspartic acid. An *E. coli* endonuclease (*nuc*) and similar proteins appear to be PLD homologues but possess only one of these motifs.

The profile contained here represents only the putative active site regions, since an accurate multiple alignment of the repeat units has not been achieved.

Number of members: 139

[1]

Medline: 96303814

A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues.

- 5 Ponting CP, Kerr ID;
Protein Sci 1996;5:914-922.
[2]Medline: 96334293

A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins.

- 10 Koonin EV;
Trends Biochem Sci 1996;21:242-243.
[3]Medline: 94327597

Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from *Ricinus communis* L.

- 15 Wang X, Xu L, Zheng L;
J Biol Chem 1994;269:20312-20317.
[4]Medline: 97386825

Regulation of eukaryotic phosphatidylinositol-specific phospholipase C and phospholipase D.

- 20 Singer WD, Brown HA, Sternweis PC;
Annu Rev Biochem 1997;66:475-509.

- 25 416. (PMI type1) Phosphomannose isomerase type I signatures
Phosphomannose isomerase (EC 5.3.1.8) (PMI) [1,2] is the enzyme that catalyzes the interconversion of mannose-6-phosphate and fructose-6-phosphate. In eukaryotes, it is involved in the synthesis of GDP-mannose which is a constituent of N- and O-linked glycans as well as GPI anchors. In prokaryotes,
30 it is involved in a variety of pathways including capsular polysaccharide biosynthesis and D-mannose metabolism.
Three classes of PMI have been defined on the basis of sequence similarities [1]. The first class comprises all known eukaryotic PMI as well as the enzyme

encoded by the *manA* gene in enterobacteria such as *Escherichia coli*. Class I PMI's are proteins of about 42 to 50 Kd which bind a zinc ion essential for their activity.

As signature patterns for class I PMI, two conserved regions were selected. The first one is located in the N-terminal section of these proteins, the second in the C-terminal half. Both patterns contain a residue involved [3] in the binding of the zinc ion.

Consensus pattern: Y-x-D-x-N-H-K-P-E [E is a zinc ligand]

-Consensus pattern: H-A-Y-[LIVM]-x-G-x(2)-[LIVM]-E-x-M-A-x-S-D-N-x-[LIVM]-R-A-G-x-T-P-K [H is a zinc ligand]

[1] Proudfoot A.E.I., Turcatti G., Wells T.N.C., Payton M.A., Smith D.J. *Eur. J. Biochem.* 219:415-423(1994).

[2] Coulin F., Magnenat E., Proudfoot A.E.I., Payton M.A., Scully P., Wells T.N.C. *Biochemistry* 32:14139-14144(1993).

[3] Cleasby A., Wonacott A., Skarzynski T., Hubbard R.E., Davies G.J., Proudfoot A.E.I., Bernard A.R., Payton M.A., Wells T.N.C. *Nat. Struct. Biol.* 3:470-479(1996).

417. (PNP UDP 1) Purine and other phosphorylases family 1 signature

The following phosphorylases belongs to the same family:

- Purine nucleoside phosphorylase (EC 2.4.2.1) (PNP) from most bacteria (gene *deoD*). This enzyme catalyzes the cleavage of guanosine or inosine to respective bases and sugar-1-phosphate molecules [1].
- Uridine phosphorylase (EC 2.4.2.3) (UdRPase) from bacteria (gene *udp*) and mammals. Catalyzes the cleavage of uridine into uracil and ribose-1-phosphate. The products of the reaction are used either as carbon and energy sources or in the rescue of pyrimidine bases for nucleotide synthesis [2].
- 5'-methylthioadenosine phosphorylase (EC 2.4.2.28) (MTA phosphorylase) from *Sulfolobus solfataricus* [3].

As a signature pattern, a conserved region was selected in the central part of

these enzymes.

Consensus pattern: [GST]-x-G-[LIVM]-G-x-[PA]-S-x-[GSTA]-I-x(3)-E-L

-Note: it should be noted that mammalian and some bacterial PNP as well as eukaryotic
5 MTA phosphorylase belong to a different family of phosphorylases (see <PDOC00954>).

[1] Takechara M., Ling F., Izawa S., Inoue Y., Kimura A. Biosci. Biotechnol. Biochem.
59:1987-1990(1995).

[2] Watanabe S.-I., Hino A., Wada K., Eliason J.F., Uchida T. J. Biol. Chem. 270:12191-
10 12196(1995).

[3] Cacciapuoti G., Porcelli M., Bertoldo C., De Rosa M., Zappia V. J. Biol. Chem.
269:24762-24769(1994).

15 418. (PP2C) Protein phosphatase 2C signature

Protein phosphatase 2C (PP2C) is one of the four major classes of mammalian
serine/threonine specific protein phosphatases (EC 3.1.3.16). PP2C [1] is a
monomeric enzyme of about 42 Kd which shows broad substrate specificity and
is dependent on divalent cations (mainly manganese and magnesium) for its
20 activity. Its exact physiological role is still unclear. Three isozymes are
currently known in mammals: PP2C-alpha, -beta and -gamma. In yeast, there are
at least four PP2C homologs: phosphatase PTC1 [2] which has weak tyrosine
phosphatase activity in addition to its activity on serines, phosphatases PTC2
and PTC3, and hypothetical protein YBR125c. Isozymes of PP2C are also known
25 from *Arabidopsis thaliana* (ABI1, PPH1), *Caenorhabditis elegans* (FEM-2,
F42G9.1, T23F11.1), *Leishmania chagasi* and *Paramecium tetraurelia*.
In *Arabidopsis thaliana*, the kinase associated protein phosphatase (KAPP) [3]
is an enzyme that dephosphorylates the Ser/Thr receptor-like kinase RLK5 and
which contains a C-terminal PP2C domain.

30 PP2C does not seem to be evolutionary related to the main family of serine/
threonine phosphatases: PP1, PP2A and PP2B. However, it is significantly
similar to the catalytic subunit of pyruvate dehydrogenase phosphatase
(EC 3.1.3.43) (PDPC) [4], which catalyzes dephosphorylation and concomitant

reactivation of the alpha subunit of the E1 component of the pyruvate dehydrogenase complex. PDPC is a mitochondrial enzyme and, like PP2C, is magnesium-dependent.

As a signature pattern, the best conserved region was selected which is located in the N-terminal part and contains a perfectly conserved tripeptide. This region includes a conserved aspartate residue involved in divalent cation binding [5].

Consensus pattern: [LIVMFY]-[LIVMFYA]-[GSAC]-[LIVM]-[FYC]-D-G-H-[GAV]

-Note: PP2C belongs [6] to a superfamily which also includes bacterial proteins such as *Bacillus spoIIE*, *rsbU* and *rsbW*, *Synechocystis* PCC 6803 *icfG* as well as a domain in fungal adenylate cyclases.

[1] Wenk J., Trompeter H.-I., Pettrich K.-G., Cohen P.T.W., Campbell D.G., Mieskes G. FEBS Lett. 297:135-138(1992).

[2] Maeda T., Tsai A.Y.M., Saito H. Mol. Cell. Biol. 13:5408-5417(1993).

[3] Stone J.M., Collinge M.A., Smith R.D., Horn M.A., Walker J.C. Science 266:793-795(1994).

[4] Lawson J.E., Niu X.-D., Browning K.S., Trong H.L., Yan J., Reed L.J. Biochemistry 32:8987-8993(1993).

[5] Das A.K., Helps N.R., Cohen P.T.W., Barford D. EMBO J. 24:6798-6809(1996).

[6] Bork P., Brown N.P., Hegyi H., Schultz J. Protein Sci. 5:1421-1425(1996).

419. (PPTA) Protein prenyltransferases alpha subunit repeat signature
Protein prenyltransferases catalyze the transfer of an isoprenyl moiety to a cysteine four residues from the C-terminus of several proteins. They are heterodimeric enzymes consisting of alpha and beta subunits. The alpha subunit is thought to participate in a stable complex with the isoprenyl substrate; the beta subunit binds the peptide substrate. Distinct protein prenyltransferases might share a common alpha subunit. Both the alpha and beta subunit show repetitive sequence motifs [1]. These repeats have distinct structural and functional implications and are unrelated to each other. Known

protein prenyltransferase alpha subunits are:

- Mammalian protein farnesyltransferase alpha subunit.
- Yeast protein RAM2, a protein farnesyltransferase alpha subunit.
- Yeast protein BET4, a protein geranylgeranyltransferase alpha subunit.

- 5 The conserved domain of the alpha subunit consists of about 34 amino acids and is repeated five times. It contains an invariant tryptophan possibly involved in heterodimerization with the conserved phenylalanines in the repeated domains of the beta subunits, via hydrophobic bonds. The signature pattern for this domain is centered on the invariant tryptophan.

10

Consensus pattern: [PSIAV]-x-[NDFV]-[NEQIY]-x-[LIVMAGP]-W-[NQSTHF]-[FYHQ]-[LIVMR]

[1] Boguski M.S., Murray A.W., Powers S. New Biol. 4:408-411(1992).

15

420. (PR55) Protein phosphatase 2A regulatory subunit PR55 signatures

Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric

20

enzyme that consists of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit (PR65), also called subunit A; this complex then associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme [1]. One of the forms of the variable subunit is

25

a 55 Kd protein (PR55) which is highly conserved in mammals - where three isoforms are known to exist -, Drosophila and yeast (gene CDC55). This subunit could perform a substrate recognition function or be responsible for targeting the enzyme complex to the appropriate subcellular compartment.

As signature patterns, two perfectly conserved sequences of 15

30

residues were selected; one located in the N-terminal region, the other in the center of the protein.

Consensus pattern: E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N

Consensus pattern: N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D

[1] Mayer-Jaekel R., Hemmings B.A. Trends Cell Biol. 4:287-291(1994).

5

421. N-(5'phosphoribosyl)anthranilate (PRA) isomerase

[1] Wilmanns M, Priestle JP, Niermann T, Jansonius JN;

J Mol Biol 1992;223:477-507.

10

422. (PRK) Phosphoribulokinase signature

Phosphoribulokinase (EC 2.7.1.19) (PRK) [1,2] is one of the enzymes specific to the Calvin's reductive pentose phosphate cycle which is the major route by which carbon dioxide is assimilated and reduced by autotrophic organisms. PRK catalyzes the ATP-dependent phosphorylation of ribulose 5-phosphate into ribulose 1,5-bisphosphate which is the substrate for RubisCO.

PRK's of diverse origins show different properties with respect to the size of the protein, the subunit structure, or the enzymatic regulation. However an alignment of the sequences of PRK from plants, algae, photosynthetic and chemoautotrophic bacteria shows that there are a few regions of sequence similarity. As a signature pattern one of these regions was selected.

Consensus pattern: K-[LIVM]-x-R-D-x(3)-R-G-x-[ST]-x-E

25

[1] Kossmann J., Klintworth R., Bowien B. Gene 85:247-252(1989).

[2] Gibson J.L., Chen J.-H., Tower P.A., Tabita F.R. Biochemistry 29:8085-8093(1990).

30

423. (PRPP synt) Phosphoribosyl pyrophosphate synthetase signature

Phosphoribosyl pyrophosphate synthetase (EC 2.7.6.1) (PRPP synthetase) catalyzes the formation of PRPP from ATP and ribose 5-phosphate. PRPP is then used in various biosynthetic pathways, as for example in the formation of purines, pyrimidines, histidine and tryptophan. PRPP synthetase requires

inorganic phosphate and magnesium ions for its stability and activity.

In mammals, three isozymes of PRPP synthetase are found; in yeast there are at least four isozymes.

As a signature pattern for this enzyme, a very conserved region was selected that has been suggested to be involved in binding divalent cations [1]. This region contains two conserved aspartic acid residues as well as a histidine, which are all potential ligands for a cation such as magnesium.

Consensus pattern: D-[LI]-H-[SA]-x-Q-[IMST]-[QM]-G-[FY]-F-x(2)-P-[LIVMFC]-D

[1] Bower S.G., Harlow K.W., Switzer R.L., Hoven-Jensen B. J. Biol. Chem. 264:10287-10291(1989).

424. (PRTP) Herpesvirus processing and transport protein

The members of this family are associated with capsid intermediates during packaging of the virus.

Number of members: 31

[1]

Medline: 98362148

Herpes simplex virus type 1 cleavage and packaging proteins

UL15 and UL28 are associated with B but not C capsids during packaging. Yu D, Weller SK;

J Virol 1998;72:7428-7439.

425. Photosystem I psaG / psaK (PSI PSAK) proteins signature

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. It is found in the chloroplasts of plants and cyanobacteria. PSI is composed of at least 14 different subunits, two of which PSI-G (gene psaG) and PSI-K (gene psaK) are small hydrophobic proteins of about 7 to 9 Kd and evolutionary related [2]. Both seem to contain two transmembrane regions. Cyanobacteria seem to encode only for PSI-K.

As a signature pattern, the best-conserved region was selected which seems to correspond to the second transmembrane region.

-Consensus pattern: [GT]-F-x-[LIVM]-x-[DEA]-x(2)-[GA]-x-[GTA]-[SA]-x-G-H-x-[LIVM]-
5 [GA]

[1] Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).

[2] Kjærulff S., Andersen B., Nielsen V.S., Møller B.L., Økkelis J.S. J. Biol. Chem. 268:18912-18916(1993).

426. PTR2 family proton/oligopeptide symporters signatures

A family of eukaryotic and prokaryotic proteins that seem to be mainly involved in the intake of small peptides with the concomitant uptake of a proton has been recently characterized [1,2]. Proteins that belong to this family are: - Fungal peptide transporter PTR2.

- Mammalian intestine proton-dependent oligopeptide transporter PepT1.

- Mammalian kidney proton-dependent oligopeptide transporter PepT2.

- *Drosophila* opt1.

- *Arabidopsis thaliana* peptide transporters PTR2-A and PTR2-B (also known as the histidine transporting protein NTR1).

- *Arabidopsis thaliana* proton-dependent nitrate/chlorate transporter CHL1.

- *Lactococcus* proton-dependent di- and tri-peptide transporter dtpT.

- *Caenorhabditis elegans* hypothetical protein C06G8.2.

- *Caenorhabditis elegans* hypothetical protein F56F4.5.

- *Caenorhabditis elegans* hypothetical protein K04E7.2.

- *Escherichia coli* hypothetical protein ybgH.

- *Escherichia coli* hypothetical protein ydgR.

- *Escherichia coli* hypothetical protein yhiP.

- *Escherichia coli* hypothetical protein yjdL.

- *Bacillus subtilis* hypothetical protein yclF.

These integral membrane proteins are predicted to comprise twelve transmembrane regions. As signature patterns, two of the best

conserved regions were selected. The first is a region that includes the end of the second transmembrane region, a cytoplasmic loop as well as the third transmembrane region. The second pattern corresponds to the core of the fifth transmembrane region.

-Consensus pattern: [GA]-[GAS]-[LIVMFYWA]-[LIVM]-[GAS]-D-x-[LIVMFYWT]-
[LIVMFYW]-G-x(3)-[TAV]-[IV]-x(3)-[GSTAV]-x-[LIVMF]-x(3)-[GA]
-Consensus pattern: [FYT]-x(2)-[LMFY]-[FYV]-[LIVMFYWA]-x-[IVG]-N-[LIVMAG]-G-
[GSA]-[LIMF]

[1] Paulsen I.T., Skurray R.A. Trends Biochem. Sci. 19:404-404(1994).

[2] Steiner H.-Y., Naider F., Becker J.M. Mol. Microbiol. 16:825-834(1995).

427. Pumilio-family RNA binding domains (aka PUM-HD, Pumilio homology domain)

Puf domains are necessary and sufficient for sequence specific RNA binding in fly Pumilio and worm FBF-1 and FBF-2. Both proteins function as translational repressors in early embryonic development by binding sequences in the 3' UTR of target mRNAs (e.g. the nanos response element (NRE) in fly Hunchback mRNA, or the point mutation element (PME) in worm fem-3 mRNA). Other proteins that contain Puf domains are also plausible RNA binding proteins. JSN1_YEAST, for instance, appears to also contain a single RRM domain by HMM analysis.

Puf domains usually occur as a tandem repeat of 8 domains. The Pfam model does not necessarily recognize all 8 domains in all sequences; some sequences appear to have 5 or 6 domains on initial analysis, but further analysis suggests the presence of additional divergent domains.

[1] Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP. Nature 1997;390:477-484. [2] Zamore PD, Williamson JR, Lehmann R. RNA 1997;3:1421-1433.

428. PWWP domain. The PWWP domain is named after a conserved Pro-Trp-Trp-Pro motif. The function of the domain is currently unknown. Number of members: 19

[1] Medline: 98282232. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Stec I, Wright TJ, van Ommen GJB, de Boer PAJ, van Haeringen A, Moorman AFM, Altherr MR, den Dunnen JT; Hum Mol Genet 1998;7:1071-1082.

429. PX domain

Eukaryotic domain of unknown function present in phox proteins, PLD isoforms, a PI3K isoform.

Number of members: 71

[1]

Medline: 97084820

Novel domains in NADPH oxidase subunits, sorting nexins, and

PtdIns 3-kinases: binding partners of SH3 domains?

Ponting CP;

Protein Sci 1996;5:2353-2357.

430. ParA family ATPase

[1]

Medline: 91141297

A family of ATPases involved in active partitioning of diverse bacterial plasmids.

Motallebi-Veshareh M, Rouch DA, Thomas CM;

Mol Microbiol 1990;4:1455-1463.

Number of members: 122

431. (Parvo coat) Parvovirus coat protein. 72 members.

432. Pectinesterase signatures

Pectinesterase (EC 3.1.1.11) (pectin methylesterase) catalyzes the hydrolysis of pectin into pectate and methanol. In plants, it plays an important role in cell wall metabolism during fruit ripening. In plant bacterial pathogens such as *Erwinia carotovora* and in fungal pathogens such as *Aspergillus niger*, pectinesterase is involved in maceration and soft-rotting of plant tissue. Prokaryotic and eukaryotic pectinesterases share a few regions of sequence similarity [1,2,3]. two of these regions were selected as signature patterns. The first is based on a region in the N-terminal section of these enzymes; it contains a conserved tyrosine which may play a role in the catalytic mechanism [3]. The second pattern corresponds to the best conserved region, an octapeptide located in the central part of these enzymes.

-Consensus pattern: [GSTNP]-x(6)-[FYVHR]-[IVN]-[KEP]-x-G-[STIVKRO]-Y-[DNQKRMV]-[EP]-x(3)-[LIMVA]

-Consensus pattern: [IV]-x-G-[STAD]-[LIVT]-D-[FYI]-[IV]-[FSN]-G

[1] Ray J., Knapp J., Grierson D., Bird C., Schuch W. Eur. J. Biochem. 174:119-124(1988).

[2] Plastow G.S. Mol. Microbiol. 2:247-254(1988).

[3] Markovic O., Joernvall H. Protein Sci. 1:1288-1292(1992).

433. Pentapeptide repeats (8 copies)

These repeats are found in many cyanobacterial proteins.

The repeats were first identified in hglK [1]. The function of these repeats is unknown.

The structure of this repeat has been predicted to be a beta-helix [2].

The repeat can be approximately described as A(D/N)LXX, where

X can be any amino acid. Number of members: 75

[1]

Medline: 96062225

The hglK gene is required for localization of

- 5 heterocyst-specific glycolipids in the cyanobacterium
Anabaena sp. strain PCC 7120.

Black K, Buikema WJ, Haselkorn R;
J Bacteriol 1995;177:6440-6448.

[2]Medline: 98318059

- 10 Structure and distribution of pentapeptide repeats in
bacteria.

Bateman A, Murzin A, Teichmann SA;
Protein Sci 1998;7:1477-1480.

[3]Medline: 98316713

- 15 Characterisation of an Arabidopsis cDNA encoding a thylakoid
lumen protein related to a novel 'pentapeptide repeat' family
of proteins.

Kieselbach T, Mant A, Robinson C, Schroder WP;
FEBS Lett 1998;428:241-244.

20

434. Polypeptide deformylase

[1]

Medline: 97002011

- 25 A new subclass of the zinc metalloproteases superfamily
revealed by the solution structure of peptide deformylase.

Meinzel T, Blanquet S, Dardel F;
J Mol Biol 1996;262:375-386.

[2]Medline: 98332750

- 30 Solution structure of nickel-peptide deformylase.

Dardel F, Ragusa S, Lazennec C, Blanquet S, Meinzel T;
J Mol Biol 1998;280:501-513.

Number of members: 21

435. Peptidyl-tRNA hydrolase signatures

Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH) is a bacterial enzyme that cleaves
 5 peptidyl-tRNA or N-acyl-aminoacyl-tRNA to yield free peptides or N-acyl-amino
 acids and tRNA. The natural substrate for this enzyme may be peptidyl-tRNA
 which drop off the ribosome during protein synthesis [1,2]. Bacterial PTH has
 been found [2,3] to be evolutionary related to yeast hypothetical protein
 YHR189w.

10 PTH and YHR189w are proteins of about 200 amino acid residues. As signature
 patterns, two conserved regions were selected that each contain an histidine.
 The first of these regions is located in the N-terminal section, the other in
 the central part.

15 -Consensus pattern: [FY]-x(2)-T-R-H-N-x-G-x(2)-[LIVMFA](2)-[DE]

-Consensus pattern: [GS]-x(3)-H-N-G-[LIVM]-[KR]-[DNS]-[LIVMT]

[1] Garcia-Villegas M.R., De La Vega F.M., Galindo J.M., Segura M., Buckingham R.H.,
 Guarneros G. EMBO J. 10:3549-3555(1991).

20 [2] De La Vega F.M., Galindo J.M., Old I.G., Guarneros G. Gene 169:97-100(1996).

[3] Ouzounis C., Bork P., Casari G., Sander C. Protein Sci. 4:2424-2428(1995).

436. (Peptidase M17) Cytosol aminopeptidase signature

25 Cytosol aminopeptidase is a eukaryotic cytosolic zinc-dependent exopeptidase
 that catalyzes the removal of unsubstituted amino-acid residues from the
 N-terminus of proteins. This enzyme is often known as leucine aminopeptidase
 (EC 3.4.11.1) (LAP) but has been shown [1] to be identical with prolyl
 aminopeptidase (EC 3.4.11.5). Cytosol aminopeptidase is a hexamer of identical
 30 chains, each of which binds two zinc ions.

Cytosol aminopeptidase is highly similar to *Escherichia coli* pepA, a manganese
 dependent aminopeptidase. Residues involved in zinc ion-binding [2] in the
 mammalian enzyme are absolutely conserved in pepA where they presumably bind

manganese.

A cytosol aminopeptidase from *Rickettsia prowazekii* [3] and one from *Arabidopsis thaliana* also belong to this family.

As a signature pattern for these enzymes, a perfectly conserved

- 5 octapeptide was selected which contains two residues involved in binding metal ions: an aspartate and a glutamate.

-Consensus pattern: N-T-D-A-E-G-R-L [The D and the E are zinc/manganese ligands]

-Note: these proteins belong to family M17 in the classification of peptidases [4,E1].

10

[1] Matsushima M., Takahashi T., Ichinose M., Miki K., Kurokawa K., Takahashi K.
Biochem. Biophys. Res. Commun. 178:1459-1464(1991).

[2] Burley S.K., David P.R., Sweet R.M., Taylor A., Lipscomb W.N. J. Mol. Biol. 224:113-140(1992).

15

[3] Wood D.O., Solomon M.J., Speed R.R. J. Bacteriol. 175:159-165(1993).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

437. Assemblin (Peptidase family S21)

20

[1]

Medline: 96399137

Three-dimensional structure of human cytomegalovirus
protease.

Shieh HS, Kurumbail RG, Stevens AM, Stegeman RA, Sturman EJ,

25

Pak JY, Wittwer AJ, Palmier MO, Wiegand RC, Holwerda BC,

Stallings WC;

Nature 1996;383:279-282.

Number of members: 29

30

438. Pollen proteins Ole e I family signature

The following plant pollen proteins, whose biological function is not yet known, are structurally related [1]:

397

- Olive tree pollen major allergen (Ole e 1).

- Tomato anther-specific protein LAT52. - Maize pollen-specific protein ZmC13.

These proteins are most probably secreted and consist of about 145 residues.

As shown in the following schematic representation, there are six cysteines

5 which are conserved in the sequence of these proteins. They seem to be involved in disulfide bonds.

xxxxxxCxXXXXXXXXXXXXXXXXXXXXXXXXXXXXCXXXXCXXXXXXXXXXXXXXXXXXXXCXXXXXX

*****C': conserved cysteine involved in a disulfide bond.

*': position of the pattern.

10

-Consensus pattern: [EQ]-G-x-V-Y-C-D-T-C-R [The two C's are probably involved in disulfide bonds]

[1] Villalba M., Batanero E., Lopez-Otin C., Sanchez L.M., Monsalve R.I., Gonzalez De La
15 Pena M.A., Lahoz C., Rodriguez R. Eur. J. Biochem. 216:863-869(1993).

439. Pollen allergen

This family contains allergens lol PI, PII and PIII from *Lolium perenne*.

20 Number of members: 49

[1]

Medline: 90105394

Complete primary structure of a *Lolium perenne* (perennial rye
grass) pollen allergen, Lol p III: comparison with known Lol

25 p I and II sequences.

Ansari AA, Shenbagamurthi P, Marsh DG;

Biochemistry 1989;28:8665-8670.

30 440. Porphobilinogen deaminase cofactor-binding site

Porphobilinogen deaminase (EC 4.3.1.8), or hydroxymethylbilane synthase, is an enzyme involved in the biosynthesis of porphyrins and related macrocycles. It catalyzes the assembly of four porphobilinogen (PBG) units in a head to tail

fashion to form hydroxymethylbilane.

The enzyme covalently binds a dipyrromethane cofactor to which the PBG subunits are added in a stepwise fashion. In the *Escherichia coli* enzyme (gene hemC), this cofactor has been shown [1] to be bound by the sulfur atom of a cysteine. The region around this cysteine is conserved in porphobilinogen deaminases from various prokaryotic and eukaryotic sources.

-Consensus pattern: E-R-x-[LIVMFA]-x(3)-[LIVMF]-x-G-[GSA]-C-x-[IVT]-P-[LIVMF]-[GSA] [C is the cofactor attachment site]

[1] Miller A.D., Hart G.J., Packman L.C., Battersby A.R. *Biochem. J.* 254:915-918(1988).

441. Presenilin

Mutations in presenilin-1 are a major cause of early onset Alzheimer's disease [2]. It has been found that presenilin-1 (Swiss:P49768) binds to beta-catenin in vivo [4]. This family also contains SPE proteins from *C.elegans*.

Number of members: 23

[1]

Medline: 98045995

Presenilins and Alzheimer's disease.

Kim TW, Tanzi RE;

Curr Opin Neurobiol 1997;7:683-688.

[2]Medline: 98045995

Presenilins and Alzheimer's disease.

Kim TW, Tanzi RE;

Curr Opin Neurobiol 1997;7:683-688.

[3]Medline: 98099802

Interaction of presenilins with the filamin family of actin-binding proteins.

Zhang W, Han SW, McKeel DW, Goate A, Wu JY;

J Neurosci 1998;18:914-922.

[4]Medline: 99004850

Destabilisation of beta-catenin by mutations in presenilin-1

potentiates neuronal apoptosis.

Zhang Z, Hartmann H, Do VM, Abramowski D, Sturchler-Pierrat

C, Staufenbiel M, Sommer B, van de Wetering M, Clevers H,

5 Saftig P, De Strooper B, He X, Yankner BA;

Nature 1998;395:698-702.

442. (Pribosyltran) Purine/pyrimidine phosphoribosyl transferases signature

10 Phosphoribosyltransferases (PRT) are enzymes that catalyze the synthesis of
beta-n-5'-monophosphates from phosphoribosylpyrophosphate (PRPP) and an enzyme
specific amine. A number of PRT's are involved in the biosynthesis of purine,
pyrimidine, and pyridine nucleotides, or in the salvage of purines and
pyrimidines. These enzymes are:

15 - Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT), which is involved in
purine salvage.

- Hypoxanthine-guanine or hypoxanthine phosphoribosyltransferase (EC 2.4.2.8)
(HGPRT or HPRT), which are involved in purine salvage.

20 - Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT), which is involved
in pyrimidine biosynthesis.

- Amido phosphoribosyltransferase (EC 2.4.2.14), which is involved in purine
biosynthesis.

- Xanthine-guanine phosphoribosyltransferase (EC 2.4.2.22) (XGPRT), which is
involved in purine salvage.

25 In the sequence of all these enzymes there is a small conserved region which
may be involved in the enzymatic activity and/or be part of the PRPP binding
site [1].

-Consensus pattern: [LIVMFYWCTA]-[LIVM]-[LIVMA]-[LIVMFC]-[DE]-D-[LIVMS]-
30 [LIVM]-[STAVD]-[STAR]-[GAC]-x-[STAR]

-Note: in position 11 of the pattern most of these enzymes have Gly.

[1] Hershey H.V., Taylor M.W. Gene 43:287-293(1986).

443. (Pro CA)

Prokaryotic-type carbonic anhydrases signatures

Carbonic anhydrases (EC 4.2.1.1) (CA) are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide. In *Escherichia coli*, CA (gene *cynT*) is involved in recycling carbon dioxide formed in the bicarbonate-dependent decomposition of cyanate by cyanase (gene *cynS*). By this action, it prevents the depletion of cellular bicarbonate [1]. In photosynthetic bacteria and plant chloroplast, CA is essential to inorganic carbon fixation [2]. Prokaryotic and plant chloroplast CA are structurally and evolutionary related and form a family distinct from the one which groups the many different forms of eukaryotic CA's (see <PDOC00146>). Hypothetical proteins *yadF* from *Escherichia coli* and HI1301 from *Haemophilus influenzae* also belong to this family. Two signature patterns were developed for this family of enzymes. Both patterns contain conserved residues that could be involved in binding zinc (cysteine and histidine).

-Consensus pattern: C-[SA]-D-S-R-[LIVM]-x-[AP]

-Consensus pattern: [EQ]-Y-A-[LIVM]-x(2)-[LIVM]-x(4)-[LIVMF](3)-x-G-H-x(2)-C-G

[1] Guilloton M.B., Korte J.J., Lamblin A.F., Fuchs J.A., Anderson P.M. J. Biol. Chem. 267:3731-3734(1992).

[2] Fukuzawa H., Suzuki E., Komukai Y., Miyachi S. Proc. Natl. Acad. Sci. U.S.A. 89:4437-4441(1992).

444. (Prolyl_oligopep)

Prolyl oligopeptidase family serine active site

The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (*Flavobacterium meningosepticum* and *Aeromonas hydrophila*); there is a high degree of sequence conservation between these sequences.

- *Escherichia coli* protease II (EC 3.4.21.83) (oligopeptidase B) (gene *prtB*) which cleaves peptide bonds on the C-terminal side of lysyl and arginyl residues.

- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.

- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: *STE13*) which is responsible for the proteolytic maturation of the alpha-factor precursor.

- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: *DAP2*).

- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase).

This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in *E.coli* protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern: D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4].

[1] Rawlings N.D., Polgar L., Barrett A.J. *Biochem. J.* 279:907-911(1991).

[2] Barrett A.J., Rawlings N.D.

□

[3] Polgar L., Szabo E.

□

5 [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

445. (Pterin 4a)

Pterin 4 alpha carbinolamine dehydratase

10

Pterin 4 alpha carbinolamine dehydratase is aka DCoH (dimerisation cofactor of hepatocyte nuclear factor 1-alpha).

Number of members: 11

15

[1] Cronk JD, Endrizzi JA, Alber T; Medline: 97052967 "High-resolution structures of the bifunctional enzyme and transcriptional coactivator DCoH and its complex with a product analogue." Protein Sci 1996;5:1963-1972.

20

446. (Pyridox oxidase)

Pyridoxamine 5'-phosphate oxidase signature

25

Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5) is a FMN flavoprotein involved in the de novo synthesis of pyridoxine (vitamin B6) and pyridoxal phosphate. It oxidizes pyridoxamine-5-P (PMP) and pyridoxine-5-P (PNP) to pyridoxal-5-P. The sequences of the enzyme from bacterial (genes pdxH or fprA) [1] and fungal (gene PDX3) [2] sources show that this protein has been highly conserved throughout evolution.

30

PdxH is evolutionary related [3] to one of the enzymes in the phenazine biosynthesis protein pathway, phzD (also known as phzG). As a signature pattern, a highly conserved region was selected located in the C-terminal part of these enzymes.

-Consensus pattern: [LIVF]-E-F-W-[OHG]-x(4)-R-[LIVM]-H-[DNE]-R

[1] Lam H.-M., Winkler M.E. J. Bacteriol. 174:6033-6045(1992).

[2] Loubbardi A., Karst F., Guilloton M., Marcireau C. J. Bacteriol. 177:1817-1823(1995).

[3] Pierson L.S. III, Gaffney T., Lam S., Gong F. FEMS Microbiol. Lett. 134:299-

5 307(1995).

447. (Pyrophosphatase)

Inorganic pyrophosphatase signature

10

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) [1,2] is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed principally as the product of the many biosynthetic reactions that utilize ATP. All known Ppases require the presence of divalent metal cations, with magnesium conferring the highest activity. Among other residues, a lysine has been postulated to be part or close to the active site. PPases have been sequenced from bacteria such as *Escherichia coli* (homohexamer), thermophilic bacteria *PS-3* and *Thermus thermophilus*, from the archaeobacteria *Thermoplasma acidophilum*, from fungi (homodimer), from a plant, and from bovine retina. In yeast, a mitochondrial isoform of PPase has been characterized which seems to be involved in energy production and whose activity is stimulated by uncouplers of ATP synthesis.

15

20

The sequences of PPases share some regions of similarities. As signature patterns a region was selected that contains three conserved aspartates that are involved in the binding of cations.

25

-Consensus pattern: D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC]

[The three D's bind divalent metal cations]

[1] Lahti R., Kolakowski L.F. Jr., Heinonen J., Vihinen M., Pohjanoksa K., Cooperman B.S. Biochim. Biophys. Acta 1038:338-345(1990).

30

[2] Cooperman B.S., Baykov A.A., Lahti R. Trends Biochem. Sci. 17:262-266(1992).

448. (Peptidase S26)

Signal peptidases I signatures.

Signal peptidases (SPases) [1] (aka leader peptidases) remove the signal peptides from secretory proteins. In prokaryotes three types of SPases are known: type I (gene *lepB*) which is responsible for the processing of the majority of exported pre-proteins; type II (gene *lsp*) which only process lipoproteins, and a third type involved in the processing of pili subunits. SPase I (EC 3.4.21.89) is an integral membrane protein that is anchored in the cytoplasmic membrane by one (in *B. subtilis*) or two (in *E. coli*) N-terminal transmembrane domains with the main part of the protein protruding in the periplasmic space. Two residues have been shown [2,3] to be essential for the catalytic activity of SPase I: a serine and an lysine. SPase I is evolutionary related to the yeast mitochondrial inner membrane protease subunit 1 and 2 (genes *IMP1* and *IMP2*) which catalyze the removal of signal peptides required for the targeting of proteins from the mitochondrial matrix, across the inner membrane, into the inter-membrane space [4]. In eukaryotes the removal of signal peptides is effected by an oligomeric enzymatic complex composed of at least five subunits: the signal peptidase complex (SPC). The SPC is located in the endoplasmic reticulum membrane. Two components of mammalian SPC, the 18 Kd (SPC18) and the 21 Kd (SPC21) subunits as well as the yeast SEC11 subunit have been shown [5] to share regions of sequence similarity with prokaryotic SPases I and yeast *IMP1/IMP2*. Three signature patterns have been developed for these proteins. The first signature contains the putative active site serine, the second signature contains the putative active site lysine which is not conserved in the SPC subunits, and the third signature corresponds to a conserved region of unknown biological significance which is located in the C-terminal section of all these proteins.

Consensus pattern: [GS]-x-S-M-x-[PS]-[AT]-[LF] [S is an active site residue]-

Consensus pattern: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY] [K is an active site residue]-

Consensus pattern: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG]-

[1] Dalbey R.E., von Heijne G. Trends Biochem. Sci. 17:474-478(1992).[2] Sung M., Dalbey R.E. J. Biol. Chem. 267:13154-13159(1992).[3] Black M.T. J. Bacteriol. 175:4957-4961(1993).[4] Nunnari J., Fox T.D., Walter P. Science 262:1997-2004(1993).[5] van Dijk

449. (Peptidase C1) Eukaryotic thiol (cysteine) proteases active sites. Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences). - Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2]. - Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2]. - Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium- activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain. - Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3]. - Human cathepsin O [4]. - Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide). - Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A. - House-dust mites allergens DerP1 and EurM1. - Cathepsin B-like proteinases from the worms *Caenorhabditis elegans* (genes *gcp-1*, *cpr-3*, *cpr-4*, *cpr-5* and *cpr-6*), *Schistosoma mansoni* (antigen SM31) and *Japonica* (antigen SJ31), *Haemonchus contortus* (genes AC-1 and AC-2), and *Ostertagia ostertagi* (CP-1 and CP-3). - Slime mold cysteine proteinases CP1 and CP2. - Cruzipain from *Trypanosoma cruzi* and *brucei*. - Throphozoite cysteine proteinase (TCP) from various *Plasmodium* species. - Proteases from *Leishmania mexicana*, *Theileria annulata* and *Theileria parva*. - Baculovirus cathepsin-like enzyme (v-cath). - *Drosophila* small optic lobes protein (gene *sol*), a neuronal protein that contains a calpain-like domain. - Yeast thiol protease BLH1/YCP1/LAP3. - *Caenorhabditis elegans* hypothetical protein C06G4.2, a calpain-like protein. Two bacterial peptidases are also part of this family: - Aminopeptidase C from *Lactococcus lactis* (gene *pepC*) [5]. - Thiol protease *tpr* from *Porphyromonas gingivalis*.

Three other proteins are structurally related to this family, but may have lost their proteolytic activity. - Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine. - Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see <PDOC00382>). - Plasmodium falciparum serine-repeat protein (SERA), the major blood stage antigen. This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine. The sequences around the three active site residues are well conserved and can be used as signature patterns.

Consensus pattern: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]- Note: the residue in position 4 of the pattern is almost always cysteine; the only exceptions are calpains (Leu), bleomycin hydrolase (Ser) and yeast YCP1 (Ser). -Note: the residue in position 5 of the pattern is always Gly except in papaya protease IV where it is Glu.

Consensus pattern: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH] [H is the active site residue]-

Consensus pattern: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue] - Note: these proteins belong to family C1 (papain-type) and C2 (calpains) in the classification of peptidases [7,E1].-

[1] Dufour E. Biochimie 70:1335-1342(1988).[2] Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).[3] Shi G.-P., Chapman H.A., Bhairi S.M., Deleuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).[4] Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).[5] Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).[6] Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).[7] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

450. (peptidase M24) Aminopeptidase P and proline dipeptidase signature (1).

Aminopeptidase P (EC 3.4.11.9) is the enzyme responsible for the release of any N-terminal amino acid adjacent to a proline residue. Proline dipeptidase (EC 3.4.13.9) (prolidase) splits

dipeptides with a prolyl residue in the carboxyl terminal position. Bacterial aminopeptidase P II (gene pepP) [1], proline dipeptidase (gene pepQ)[2], and human proline dipeptidase (gene PEPD) [3] are evolutionary related. These proteins are manganese metalloenzymes. Yeast hypothetical proteins YER078c and YFR006w and Mycobacterium tuberculosis hypothetical protein MtCY49.29c also belong to this family. As a signature pattern for these enzymes a conserved region that contains three histidine residues has been developed

Consensus pattern: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE]-

[1] Yoshimoto T., Tone H., Honda T., Osatomi K., Kobayashi R., Tsuru D. J. Biochem. 105:412-416(1989).[2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:6439-6439(1990).[3] Endo F., Tanoue A., Nakai H., Hata A., Indo Y., Titani K., Matsuda I. J. Biol. Chem. 264:4476-4481(1989).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

Methionine aminopeptidase signatures. (2). Methionine aminopeptidase (EC 3.4.11.18) (MAP) is responsible for the removal of the amino-terminal (initiator) methionine from nascent eukaryotic cytosolic and cytoplasmic prokaryotic proteins if the penultimate amino acid is small and uncharged. All MAP studied to date are monomeric proteins that require cobalt ions for activity. Two subfamilies of MAP enzymes are known to exist [1,2]. While being evolutionary related, they only share a limited amount of sequence similarity mostly clustered around the residues shown, in the Escherichia coli MAP [3], to be involved in cobalt-binding. The first family consists of enzymes from prokaryotes as well as eukaryotic MAP-1, while the second group is made up of archaebacterial MAP and eukaryotic MAP-2. The second subfamily also includes proteins which do not seem to be MAP, but that are clearly evolutionary related such as mouse proliferation-associated protein 1 and fission yeast curved DNA-binding protein. For each of these subfamilies, a specific signature pattern that includes residues known to be involved in cobalt-binding has been developed.

Consensus pattern: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]-[YWV]
[H is a cobalt ligand]-

Consensus pattern: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-
[DN] [The second D and the last D/N are cobalt ligands]

[1] Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E., Matthews B.W.,
Bradshaw R.A. Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).[2] Keeling P.J., Doolittle
W.F. Trends Biochem. Sci. 21:285-286(1996).[3] Roderick S.L., Mathews B.W.
Biochemistry 32:3907-3912(1993).[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-
228(1995).

451. Cytochrome P450 cysteine heme-iron ligand signature

Cytochrome P450's [1,2,3,E1] are a group of enzymes involved in the oxidative metabolism
of a high number of natural compounds (such as steroids, fatty acids, prostaglandins,
leukotrienes, etc) as well as drugs, carcinogens and mutagens. Based on sequence similarities,
P450's have been classified into about forty different families [4,5]. P450's are proteins of
400 to 530 amino acids; the only exception is *Bacillus BM-3* (CYP102) which is a protein of
1048 residues that contains a N-terminal P450 domain followed by a reductase domain.
P450's are heme proteins. A conserved cysteine residue in the C-terminal part of P450's is
involved in binding the heme iron in the fifth coordination site. From a region around this
residue, a ten residue signature was developed specific to P450's.

Consensus pattern: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD] [C is the
heme iron ligand]-

[1] Nebert D.W., Gonzalez F.J. Annu. Rev. Biochem. 56:945-993(1987).
[2] Coon M.J., Ding X., Pernecky S.J., Vaz A.D.N. FASEB J. 6:669-673(1992).
[3] Guengerich F.P. J. Biol. Chem. 266:10019-10022(1991).
[4] Nelson D.R., Kamataki T., Waxman D.J., Guengerich F.P., Estrabrook R.W., Feyereisen
R., Gonzalez F.J., Coon M.J., Gunsalus I.C., Gotoh O., Okuda K., Nebert D.W. DNA Cell
Biol. 12:1-51(1993).
[5] Degtyarenko K.N., Archakov A.I. FEBS Lett. 332:1-8(1993).

452. (Pec Lyase) Pectate lyase

This enzyme forms a right handed beta helix structure. Pectate lyase is an enzyme involved in the maceration and soft rotting of plant tissue.

[1] Yoder MD, Keen NT, Jurnak F, Science 1993;260:1503-1507.

453. (pep M24) Aminopeptidase P and proline dipeptidase signature (pep1)

Aminopeptidase P (EC [3.4.11.9](#)) is the enzyme responsible for the release of any N-terminal amino acid adjacent to a proline residue. Proline dipeptidase (EC [3.4.13.9](#)) (prolidase) splits dipeptides with a prolyl residue in the carboxyl terminal position. Bacterial aminopeptidase P II (gene pepP) [1], proline dipeptidase (gene pepQ)[2], and human proline dipeptidase (gene PEPD) [3] are evolutionary related. These proteins are manganese metalloenzymes. Yeast hypothetical proteins YER078c and YFR006w and Mycobacterium tuberculosis .hypothetical protein MTCY49.29c also belong to this family. As a signature pattern for these enzymes a conserved region was selected that contains three histidine residues.

Consensus pattern: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE]-

[1] Yoshimoto T., Tone H., Honda T., Osatomi K., Kobayashi R., Tsuru D. J. Biochem. 105:412-416(1989).

[2] Nakahigashi K., Inokuchi H. Nucleic Acids Res. 18:6439-6439(1990).

[3] Endo F., Tanoue A., Nakai H., Hata A., Indo Y., Titani K., Matsuda I. J. Biol. Chem. 264:4476-4481(1989).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

Methionine aminopeptidase signatures (pep2)

Methionine aminopeptidase (EC [3.4.11.18](#)) (MAP) is responsible for the removal of the amino-terminal (initiator) methionine from nascent eukaryotic cytosolic and cytoplasmic prokaryotic proteins if the penultimate amino acid is small and uncharged. All MAP studied to date are monomeric proteins that require cobalt ions for activity. Two subfamilies of MAP enzymes are known to exist [1,2]. While being evolutionary related, they only share a limited amount of sequence similarity mostly clustered around the residues shown, in the Escherichia coli MAP [3], to be involved in cobalt-binding. The first family consists of enzymes from

Consensus pattern: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]- [YWV]
[H is a cobalt ligand]-

Consensus pattern: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN] [The second D and the last D/N are cobalt ligands]

- [1] Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E., Matthews B.W., Bradshaw R.A. Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).
- [2] Keeling P.J., Doolittle W.F. Trends Biochem. Sci. 21:285-286(1996).
- [3] Roderick S.L., Mathews B.W. Biochemistry 32:3907-3912(1993).
- [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

454. Peroxidases signatures

Peroxidases (EC 1.11.1.-) [1] are heme-binding enzymes that carry out a variety of biosynthetic and degradative functions using hydrogen peroxide as the electron acceptor. Peroxidases are widely distributed throughout bacteria, fungi, plants, and vertebrates. In peroxidases the heme prosthetic group is protoporphyrin IX and the fifth ligand of the heme iron is a histidine (known as the proximal histidine). Another histidine residue (the distal histidine) serves as an acid-base catalyst in the reaction between hydrogen peroxide and the enzyme. The regions around these two active site residues are more or less conserved in a majority of peroxidases [2,3]. The enzymes in which one or both of these regions can be found are listed below. - Yeast cytochrome c peroxidase (EC 1.11.1.5). - Myeloperoxidase (EC 1.11.1.7) (MPO). MPO is found in granulocytes and monocytes and plays a major role in the oxygen-dependent microbicidal system of neutrophils. - Lactoperoxidase (EC 1.11.1.7) (LPO). LPO is a milk protein which acts as an antimicrobial agent. - Eosinophil peroxidase (EC 1.11.1.7) (EPO). An enzyme found in the cytoplasmic granules of eosinophils. - Thyroid

peroxidase (EC 1.11.1.8) (TPO). TPO plays a central role in the biosynthesis of thyroid hormones. It catalyzes the iodination and coupling of the hormonogenic tyrosines in thyroglobulin to yield the thyroid hormones T3 and T4. - Fungal ligninases. Ligninase catalyzes the first step in the degradation of lignin. It depolymerizes lignin by catalyzing the C(alpha)-C(beta) cleavage of the propyl side chains of lignin. - Plant peroxidases (EC 1.11.1.7). Plants express a large number of isozymes of peroxidases. Some of them play a role in cell-suberization by catalyzing the deposition of the aromatic residues of suberin on the cell wall, some are expressed as a defense response toward wounding, others are involved in the metabolism of auxin and the biosynthesis of lignin. - Prokaryotic catalase-peroxidases. Some bacterial species produce enzymes that exhibit both catalase and broad-spectrum peroxidase activities [4]. Examples of such enzymes are: catalase HP I from *Escherichia coli* (gene *katG*) and *perA* from *Bacillus stearothermophilus*.

Consensus pattern: [DET]-[LIVMTA]-x(2)-[LIVM]-[LIVMSTAG]-[SAG]-[LIVMSTAG]-H-[STA]-[LIVMFY] [H is the proximal heme-binding ligand] -

Consensus pattern: [SGATV]-x(3)-[LIVMA]-R-[LIVMA]-x-[FW]-H-x-[SAC] [H is an active site residue]-

[1] Dawson J.H. Science 240:433-439(1988).

[2] Kimura S., Ikeda-Saito M. Proteins 3:113-120(1988).

[3] Henrissat B., Saloheimo M., Lavaitte S., Knowles J.K.C. Proteins 8:251-257(1990).

[4] Welinder K.G. Biochim. Biophys. Acta 1080:215-220(1991).

455. pfkB family of carbohydrate kinases signatures

It has been shown [1,2,3] that the following carbohydrate and purine kinases are evolutionary related and can be grouped into a single family, which is known [1] as the 'pfkB family': - Fructokinase (EC 2.7.1.4) (gene *scrK*). - 6-phosphofructokinase isozyme 2 (EC 2.7.1.11) (phosphofructokinase-2) (gene *pfkB*). *pfkB* is a minor phosphofructokinase isozyme in *Escherichia coli* and is not evolutionarily related to the major isozyme (gene *pfkA*). Plants 6-phosphofructokinase also belong to this family. - Ribokinase (EC 2.7.1.15) (gene *rbsK*). - Adenosine kinase (EC 2.7.1.20) (gene *ADK*). - 2-dehydro-3-deoxygluconokinase (EC 2.7.1.45) (gene: *kdgK*). - 1-phosphofructokinase (EC 2.7.1.56) (fructose 1-phosphate kinase)

(gene fruK). - Inosine-guanosine kinase (EC 2.7.1.73) (gene gsk). - Tagatose-6-phosphate kinase (EC 2.7.1.144) (phosphotagatokinase) (gene lacC). - Escherichia coli hypothetical protein yeiC. - Escherichia coli hypothetical protein yeiL. - Escherichia coli hypothetical protein yhfQ. - Escherichia coli hypothetical protein yihV. - Bacillus subtilis hypothetical protein yxdC. - Yeast hypothetical protein YJR105w. All the above kinases are proteins of from 280 to 430 amino acid residues that share a few region of sequence similarity. Two of these regions were selected as signature patterns. The first pattern is based on a region rich in glycine which is located in the N-terminal section of these enzymes; while the second pattern is based on a conserved region in the C-terminal section.

Consensus pattern: [AG]-G-x(0,1)-[GAP]-x-N-x-[STA]-x(6)-[GS]-x(9)-G-

Consensus pattern: [DNSK]-[PSTV]-x-[SAG](2)-[GD]-D-x(3)-[SAGV]-[AG]-[LIVMFYA]-[LIVMSTAP]

[1] Wu L.-F., Reizer A., Reizer J., Cai B., Tomich J.M., Saier M.H. Jr. J. Bacteriol. 173:3117-3127(1991).

[2] Orchard L.M.D., Kornberg H.L. Proc. R. Soc. Lond., B, Biol. Sci. 242:87-90(1990).

[3] Blatch G.L., Scholle R.R., Woods D.R. Gene 95:17-23(1990).

456. Phospholipase A2 active sites signatures

Phospholipase A2 (EC 3.1.1.4) (PA2) [1,2] is an enzyme which releases fatty acids from the second carbon group of glycerol. PA2's are small and rigid proteins of 120 amino-acid residues that have four to seven disulfide bonds. PA2 binds a calcium ion which is required for activity. The side chains of two conserved residues, a histidine and an aspartic acid, participate in a 'catalytic network'. Many PA2's have been sequenced from snakes, lizards, bees and mammals. In the latter, there are at least four forms: pancreatic, membrane-associated as well as two less characterized forms. The venom of most snakes contains multiple forms of PA2. Some of them are presynaptic neurotoxins which inhibit neuromuscular transmission by blocking acetylcholine release from the nerve termini. Two different signature patterns were derived for PA2's. The first is centered on the active site histidine and contains three cysteines involved in disulfide bonds. The second is centered on the active site aspartic acid and also contains three cysteines involved in disulfide bonds.

Consensus pattern: C-C-x(2)-H-x(2)-C [H is the active site residue] This pattern will not detect some snake toxins homologous with PA2 but which have lost their catalytic activity as well as otoconin-22, a *Xenopus* protein from the aragonitic otoconia which is also unlikely to be enzymatically active.

Consensus pattern: [LIVMA]-C-{LIVMFYWPCST}-C-D-x(5)-C [D is the active site residue] The majority of functional and non-functional PA2's. Undetected sequences are bee PA2, gila monster PA2's, PA2 PL-X from habu and PA2 PA-5 from mulga.

[1] Davidson F.F., Dennis E.A. J. Mol. Evol. 31:228-238(1990).

[2] Gomez F., Vandermeers A., Vandermeers-Piret M.-C., Herzog R., Rathe J., Stievenart M., Winand J., Christophe J. Eur. J. Biochem. 186:23-33(1989).

457. Phosphorylase pyridoxal-phosphate attachment site. Phosphorylases (EC 2.4.1.1) [1] are important allosteric enzymes in carbohydrate metabolism. They catalyze the formation of glucose 1-phosphate from polyglucose such as glycogen, starch or maltodextrin. Enzymes from different sources differ in their regulatory mechanisms and their natural substrates. However, all known phosphorylases share catalytic and structural properties. They are pyridoxal-phosphate dependent enzymes; the pyridoxal-P group is attached to a lysine residue around which the sequence is highly conserved and can be used as a signature pattern to detect this class of enzymes.

Consensus pattern: E-A-[SC]-G-x-[GS]-x-M-K-x(2)-[LM]-N [K is the pyridoxal-P attachment site]-

[1] Fukui T., Shimomura S., Nakano K. Mol. Cell. Biochem. 42:129-144(1982).

458. Protein kinases signatures and profile

Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein

kinases. Two of these regions were selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; Two signature patterns were derived for that region: one specific for serine/threonine kinases and the other for tyrosine kinases. A profile was also developed which is based on the alignment in [1] and covers the entire catalytic domain.

Consensus pattern: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]. The majority of known protein kinases belong to the class detected by this pattern, but it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]. Most serine/ threonine specific protein kinases belong to this class detected by the pattern with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and *Drosophila* ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3) [D is an active site residue] ALL tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk belong to this class detected by the pattern. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses gangciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. This profile also detects receptor guanylate cyclases and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects *Arabidopsis thaliana* kinase- like protein TMKL1 which seems to have lost its catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Eukaryotic-type protein kinases have also been found in prokaryotes such as *Myxococcus xanthus* [11] and *Yersinia pseudotuberculosis*.

- [1] Hanks S.K., Hunter T. FASEB J. 9:576-596(1995).
 [2] Hunter T. Meth. Enzymol. 200:3-37(1991).
 [3] Hanks S.K., Quinn A.M. Meth. Enzymol. 200:38-62(1991).
 [4] Hanks S.K. Curr. Opin. Struct. Biol. 1:369-383(1991).
 5 [5] Hanks S.K., Quinn A.M., Hunter T. Science 241:42-52(1988).
 [6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor S.S., Sowadski J.M. Science 253:407-414(1991).
 [7] Bairoch A., Claverie J.-M. Nature 331:22(1988).
 [8] Benner S. Nature 329:21-21(1987).
 10 [9] Kirby R. J. Mol. Evol. 30:489-492(1992).
 [10] Littler E., Stuart A.D., Chee M.S. Nature 358:160-162(1992).
 [11] Munoz-Dorado J., Inouye S., Inouye M. Cell 67:995-1006(1991).

Receptor tyrosine kinase class II signature

15 A number of growth factors stimulate mitogenesis by interacting with a family of cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1]. These receptor tyrosine kinases (RTK) all share the same topology: an extracellular ligand-binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups. The prototype for class II RTK's is the insulin

20 receptor, a heterotetramer of two alpha and two beta chains linked by disulfide bonds. The alpha and beta chains are cleavage products of a precursor molecule. The alpha chain contains the ligand binding site, the beta chain transverse the membrane and contains the tyrosine protein kinase domain. The receptors currently known to belong to class II are: - Insulin receptor from vertebrates. - Insulin growth factor I receptor from mammals. - Insulin

25 receptor-related receptor (IRR), which is most probably a receptor for a peptide belonging to the insulin family. - Insect insulin-like receptors. - Molluscan insulin-related peptide(s) receptor (MIP-R). - Insulin-like peptide receptor from *Branchiostoma lanceolatum*. - The *Drosophila* developmental protein sevenless, a putative receptor for positional information required for the formation of the R7 photoreceptor cells. - The *trk* family of receptors

30 (NTRK1, NTRK2 and NTRK3), which are high affinity receptors for nerve growth factor and related neurotrophic factors (BDNF and NT-3). And the following uncharacterized receptors: - ROS. - LTK (TYK1). - EDDR1 (cck, TRKE, RTK6). - NTRK3 (Tyro10, TKT). - A sponge putative receptor tyrosine kinase. While only the insulin and the insulin growth factor I

receptors are known to exist in the tetrameric conformation specific to class II RTK's, all the above proteins share extensive homologies in their kinase domain, especially around the putative site of autophosphorylation. Hence, a signature pattern was developed for this class of RTK's, which includes the tyrosine residue, itself probably autophosphorylated.

Consensus pattern: [DN]-[LIV]-Y-x(3)-Y-Y-R [The second Y is the autophosphorylation site]

[1] Yarden Y., Ullrich A. Annu. Rev. Biochem. 57:443-478(1988).

Receptor tyrosine kinase class III signature

A number of growth factors stimulate mitogenesis by interacting with a family of cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1]. These receptor tyrosine kinases (RTK) all share the same topology: an extracellular ligand-binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups. The class III RTK's are characterized by the presence of five to seven immunoglobulin-like domains [2] in their extracellular section. Their kinase domain differs from that of other RTK's by the insertion of a stretch of 70 to 100 hydrophilic residues in the middle of this domain. The receptors currently known to belong to class III are: - Platelet-derived growth factor receptor (PDGF-R). PDGF-R exists as a homo- or heterodimer of two related chains: alpha and beta [3]. - Macrophage colony stimulating factor receptor (CSF-1-R) (also known as the *fms* oncogene). - Stem cell factor (mast cell growth factor) receptor (also known as the *kit* oncogene). - Vascular endothelial growth factor (VEGF) receptors Flt-1 and Flk-1/KDR [4]. - Fl cytokine receptor Flk-2/Flt-3 [5]. - The putative receptor Flt-4 [7]. a signature pattern Was developed for this class of RTK's which is based on a conserved region in the kinase domain.

Consensus pattern: G-x-H-x-N-[LIVM]-V-N-L-L-G-A-C-T-

[1] Yarden Y., Ullrich A. Annu. Rev. Biochem. 57:443-478(1988).

[2] Hunkapiller T., Hood L. Adv. Immunol. 44:1-63(1989).

[3] Lee K.-H., Bowen-Pope D.F., Reed R.R. Mol. Cell. Biol. 10:2237-2246(1990).

- [4] Terman B.I., Dougher-Vermazen M., Carrion M.E., Dimitrov D., Armellino D.C., Gospodarowicz D., Boehlen P. *Biochem. Biophys. Res. Commun.* 187:1579-1586(1992).
- [5] Lyman S.D., James L., Vanden Bos T., de Vries P., Brasel K., Gliniak B., Hollingsworth L.T., Picha K.S., McKenna H.J., Splett R.R. *Cell* 75:1157-1167(1993).
- 5 [6] Galland F., Karamysheva A., Pebusque M.J., Borg J.P., Rottapel R., Dubreuil P., Rosnet O., Birnbaum D. *Oncogene* 8:1233-1240(1993).

Receptor tyrosine kinase class V signatures

A number of growth factors stimulate mitogenesis by interacting with a family of cell surface receptors which possess an intrinsic, ligand-sensitive, protein tyrosine kinase activity [1].

These receptor tyrosine kinases (RTK) all share the same topology: an extracellular ligand-binding domain, a single transmembrane region and a cytoplasmic kinase domain. However they can be classified into at least five groups on the basis of sequence similarities. The extracellular domain of class V RTK's consist of a region of about 300 amino acids, amongst which 16 conserved cysteines probably involved in disulfide bonds; this region is followed by two copies of a fibronectin type III domain. The ligands for these receptors are proteins of about 200 to 300 residues collectively known as Ephrins. The receptors currently known to belong to class V are [2,3, E1]: - EPHA1 (Eph-1; Esk). - EPHA2 (Eck; Mpk-5; Sek-2). - EPHA3 (Etk-1; Hek; Mek4; Tyro4; Rek4; Cek4). - EPHA4 (Sek; Hek8; Mpk-3; Cek8). - EPHA5 (Ehk-1; Hek7; Bsk; Cek7). - EPHA6 (Ehk-2). - EPHA7 (Ehk-3; Hek11; Mdk-1; Ebk). - EPHA8 (Eek). - EPHB1 (Eph-2; Elk; Net). - EPHB2 (Eph-3; Hek5; Drt; Erk; Nuk; Sek-3; Cek5; Qek5). - EPHB3 (Hek-2; Mdk-5). - EPHB4 (Htk; Mdk-2; Myk-1). - EPHB5 (Cek9). The EPHA subtype receptors bind to GPI-anchored ephrins while the EPHB subtype receptors bind to type-I membrane ephrins. Two signature patterns were developed for this class of RTK's, which each include some of the conserved cysteine residues.

Consensus pattern: F-x-[DN]-x-[GAW]-[GA]-C-[LIVM]-[SA]-[LIVM](2)-[SA]-[LV]-[KRHQ]-[LIVA]-x(3)-[KR]-C-[PSAW] [The two C's are probably involved in disulfide bonds]

Consensus pattern: C-x(2)-[DE]-G-[DEQ]-W-x(2,3)-[PAQ]-[LIVMT]-[GT]-x-C-x-C- x(2)-G-[HFY]-[EQ] [The three C's are probably involved in disulfide bonds]

- [1] Yarden Y., Ullrich A. *Annu. Rev. Biochem.* 57:443-478(1988).

[2] Sajjadi F.G., Pasquale E.B., Subramani S. New Biol. 3:769-778(1991).

[3] Wicks I.P., Wilkinson D., Salvaris E., Boyd A.W. Proc. Natl. Acad. Sci. U.S.A. 89:1611-1615(1992).

5

459. Protein kinase C terminal domain

460. Plant thionins signature

10

Thionins are small, basic, plant proteins generally toxic to animal cells [1]. They seem to exert their toxic effect at the level of the cell membrane but their exact function is not known. They consist of a polypeptide chain of forty five to fifty amino acids with three to four internal disulfide bonds. They are found in seeds but also in the cell wall of leaves [2]. Thionins are processed from larger precursor proteins [3]. Crambin [4], a hydrophobic plant seed protein, also belongs to this family. The pattern to detect this family of proteins includes three of the six cysteine residues involved in disulfide bonds. +-----+ | +-----

15

-----+ ||| || xxCCxxxxxxxxCxxxxxxxxCxxxCxxCxxxxxCxxxxxxxx

***** ||| +-----+'C': conserved cysteine involved in a disulfide bond.'*':

position of the pattern.

20

Consensus pattern: C-C-x(5)-R-x(2)-[FY]-x(2)-C [The three C's are involved in disulfide bonds] The proteins from the gamma-thionin family are not related to the above proteins and are described in a separate section.

25

[1] Vernon L.P., Evett G.E., Zeikus R.D., Gray W.R. Arch. Biochem. Biophys. 238:18-29(1985).

[2] Bohlmann H., Clausen S., Behnke S., Giese H., Hiller C., Reimann-Phillip U., Schrader G., Barkholt V., Apel K. EMBO J. 7:1559-1565(1988).

[3] Bohlmann H., Apel K. Mol. Gen. Genet. 207:446-454(1987).

30

[4] Teeter M.M., Mazer J.A., L'Italien J.J. Biochemistry 20:5437-5443(1981).

461. Polyprenyl synthetases signatures

A variety of isoprenoid compounds are synthesized by various organisms. For example in eukaryotes the isoprenoid biosynthetic pathway is responsible for the synthesis of a variety of end products including cholesterol, dolichol, ubiquinone or coenzyme Q. In bacteria this pathway leads to the synthesis of isopentenyl tRNA, isoprenoid quinones, and sugar carrier lipids. Among the enzymes that participate in that pathway, are a number of polyprenyl synthetase enzymes which catalyze a 1'4-condensation between 5 carbon isoprene units. Currently the sequence of some of these enzymes is known: - Eukaryotic farnesyl pyrophosphate synthetase (FPP synthetase) (EC 2.5.1.1 / EC 2.5.1.10) which catalyzes the sequential condensation of isopentenyl pyrophosphate (IPP) with dimethylallyl pyrophosphate (DMAPP), and then with the resultant geranyl pyrophosphate to form farnesyl pyrophosphate. FPP synthetase is a cytoplasmic dimeric enzyme. - Prokaryotic farnesyl pyrophosphate synthetase (gene *ispA*). - Prokaryotic octaprenyl diphosphate synthase (gene *ispB*). - Prokaryotic heptaprenyl diphosphate synthase (EC 2.5.1.30). - Eukaryotic geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (EC 2.5.1.1 / EC 2.5.1.10 / EC 2.5.1.29) which catalyzes the sequential addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate. In plants GGPP synthase is a chloroplast enzyme involved in the biosynthesis of terpenoids; in fungi, such as *Neurospora crassa* (gene *al-3*), this enzyme is involved in the biosynthesis of carotenoids. - Prokaryotic GGPP synthetase, which are involved in the biosynthesis of carotenoids (gene *crtE*). Such an enzyme is also encoded in the cyanelle genome of *Cyanophora paradoxa*. - Eukaryotic hexaprenyl pyrophosphate synthetase, which is involved in the biosynthesis of coenzyme Q and which catalyzes the formation of all trans- polyprenyl pyrophosphates generally ranging in length of between 6 and 10 isoprene units depending on the species. HP synthetase is a mitochondrial membrane-associated enzyme. It has been shown [1 to 5] that all the above enzymes share some regions of sequence similarity. Two of these regions are rich in aspartic-acid residues and could be involved in the catalytic mechanism and/or the binding of the substrates. signature patterns were developed for both regions. Possible additional members of this family of proteins are: - *Bacillus subtilis* spore germination protein C3 (gene *gerC3*). Both proteins are most probably also enzymes involved in isoprenoid metabolism [6].

Consensus pattern: [LIVM](2)-x-D-D-x(2,4)-D-x(4)-R-R-[GH]-

Consensus pattern: [LIVMFY]-G-x(2)-[FYL]-Q-[LIVM]-x-D-D-[LIVMFY]-x-[DNG]

- [1] Ashby M.N., Edwards P.A. J. Biol. Chem. 265:13157-13164(1990).
[2] Fujisaki S., Hara H., Nishimura Y., Horiuchi K., Nishino T. J. Biochem. 108:995-1000(1990).
[3] Carattoli A., Romano N., Ballario P., Morelli G., Macino G. J. Biol. Chem. 266:5854-5859(1991).
[4] Kuntz M., Roemer S., Suire C., Hugueney P., Weil J.H., Schantz R., Camara B. Plant J. 2:25-34(1992).
[5] Math S.K., Hearst J.E., Poulter C.D. Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764(1992).
[6] Bairoch A. Unpublished observations (1993).

462. Potato inhibitor I family signature

The potato inhibitor I family is one of the numerous families of serine proteinase inhibitors. Members of this protein family are found in plants; in the seeds of barley or beans [1,2,3], and in potato or tomato leaves where they accumulate in response to mechanical damage [4,5]. An inhibitor belonging to this family is also found in leech [6]. It is interesting to note that, currently, this is the only proteinase inhibitor family to be found both in plant and animal kingdoms. Structurally these inhibitors are small (60 to 90 residues) and in contrast with other families of protease inhibitors, they lack disulfide bonds. They have a single inhibitory site. The consensus pattern includes three out of the four residues conserved in all members of this family and is located in the N-terminal half.

Consensus pattern: [FYW]-P-[EQH]-[LIV](2)-G-x(2)-[STAGV]-x(2)-A- Barley subtilisin-chymotrypsin inhibitor-2b has Glu instead of Gly. There is a trypsin inhibitor from the cucurbitaceae *Momordica charantia* [7], which is said to belong to the potato inhibitor I family but which shows only a very weak similarity with the other members of this family.

- [1] Svendsen I., Hejgaard J., Chavan J.K. Carlsberg Res. Commun. 49:493-502(1984).
[2] Svendsen I., Boisen S., Hejgaard J. Carlsberg Res. Commun. 47:45-53(1982).
[3] Nozawa H., Yamagata H., Aizono Y., Yoshikawa M., Iwasaki T. J. Biochem. 106:1003-1008(1989).
[4] Cleveland T.E., Thornburg R.W., Ryan C.A. Plant Mol. Biol. 8:199-207(1987).

- [5] Lee J.S., Brown W.E., Graham J.S., Pearce G., Fox E.A., Dreher T.W., Ahern K.G., Pearson G.D., Ryan C.A. Proc. Natl. Acad. Sci. U.S.A. 83:7277-7281(1986).
- [6] Seemuller U., Eulitz M., Fritz H., Strobl A. Hoppe-Seyler's Z. Physiol. Chem. 361:1841-1846(1980).
- 5 [7] Zeng F.-Y., Qian R.-Q., Wang Y. FEBS Lett. 234:35-38(1988).

463. (pp binding) Phosphopantetheine attachment site

Phosphopantetheine (or pantetheine 4' phosphate) is the prosthetic group of acyl carrier proteins (ACP) in some multienzyme complexes where it serves as a 'swinging arm' for the attachment of activated fatty acid and amino-acid groups [1]. Phosphopantetheine is attached to a serine residue in these proteins [2]. ACP proteins or domains have been found in various enzyme systems which are listed below (references are only provided for recently determined sequences). - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to the different enzymatic activities; ACP is one of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the ACP domain is located in the N-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional enzyme; the ACP domain is located between the beta-ketoacyl reductase domain and the C-terminal thioesterase domain [3]. - Polyketide antibiotics synthase enzyme systems. Polyketides are secondary metabolites produced from simple fatty acids, by microorganisms and plants. ACP is one of the polypeptidic components involved in the biosynthesis of *Streptomyces* polyketide antibiotics actinorhodin, curamycin, granatacin, monensin, oxytetracycline and tetracenomycin C. - *Bacillus subtilis* putative polyketide synthases pksK, pksL and pksM which respectively contain three, five and one ACP domains. - The multifunctional 6-methylsalicylic acid synthase (MSAS) from *Penicillium patulum*. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which contains an ACP domain in the C-terminal extremity. - Multifunctional mycocerosic acid synthase (gene mas) from *Mycobacterium bovis*. - Gramicidin S synthetase I (gene grsA) from *Bacillus brevis*. This enzyme catalyzes the first step in the biosynthesis of the cyclic antibiotic gramicidin S. - Tyrocidine synthetase I (gene tycA) from *Bacillus brevis*. The reaction carried out by tycA is identical to that catalyzed by grsA - Gramicidin S synthetase II (gene grsB) from *Bacillus brevis*. This enzyme is a

multifunctional protein that activates and polymerizes proline, valine, ornithine and leucine. GrsB contains four ACP domains. - Erythronolide synthase proteins 1, 2 and 3 from *Saccharopolyspora erythraea* which is involved in the biosynthesis of the polyketide antibiotic erythromycin. Each of these proteins contain two ACP domains. - Conidial green pigment synthase from *Aspergillus nidulans*. - ACV synthetase from various fungi. This enzyme catalyzes the first step in the biosynthesis of penicillin and cephalosporin. It contains three ACP domains. - Enterobactin synthetase component F (gene *entF*) from *Escherichia coli*. This enzyme is involved in the ATP-dependent activation of serine during enterobactin (enterochelin) biosynthesis. - Cyclic peptide antibiotic surfactin synthase subunits 1, 2 and 3 from *Bacillus subtilis*. Subunits 1 and 2 contains three related domains while subunit 3 only contains a single domain. - HC-toxin synthetase (gene *HTS1*) from *Cochliobolus carbonum*. This enzyme synthesizes HC-toxin, a cyclic tetrapeptide. *HTS1* contains four ACP domains. - Fungal mitochondrial ACP [9], which is part of the respiratory chain NADH dehydrogenase (complex I). - *Rhizobium* nodulation protein *nodF*, which probably acts as an ACP in the synthesis of the nodulation Nod factor fatty acyl chain. The sequence around the phosphopantetheine attachment site is conserved in all these proteins and can be used as a signature pattern. A profile was also developed that spans the complete ACP-like domain.

Consensus pattern: [DEQGSTALMKRH]-[LIVMFYSTAC]-[GNQ]-[LIVMFYAG]-[DNEKHS]-S-[LIVMST]-{PCFY}-[STAGCPQLIVMF]-[LIVMATN]-[DENQGTAKRHLM]-[LIVMWSTA]-[LIVGSTACR]-x(2)-[LIVMFA] [S is the pantetheine attachment site]

[1] Concise Encyclopedia Biochemistry, Second Edition, Walter de Gruyter, Berlin New-York (1988).

[2] Pugh E.L., Wakil S.J. J. Biol. Chem. 240:4727-4733(1965).

[3] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991).

[6] Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G., Galizzi A., Albertini A.M. Gene 130:65-71(1993).

[9] Sackmann U., Zensen R., Rohlen D., Jahnke U., Weiss H. Eur. J. Biochem. 200:463-469(1991).

464. (Prenyltrans) Terpene synthases signature

The following enzymes catalyze mechanistically related reactions which involve the highly complex cyclic rearrangement of squalene or its 2,3 oxide: - Lanosterol synthase (EC 5.4.99.7) (oxidosqualene--lanosterol cyclase), which catalyzes the cyclization of (S)-2,3-epoxysqualene to lanosterol, the initial precursor of cholesterol, steroid hormones and vitamin D in vertebrates and of ergosterol in fungi (gene ERG7). - Cycloartenol synthase (EC 5.4.99.8) (2,3-epoxysqualene--cycloartenol cyclase), a plant enzyme that catalyzes the cyclization of (S)-2,3-epoxysqualene to cycloartenol. - Hopene synthase (EC 5.4.99.-) (squalene--hopene cyclase), a bacterial enzyme that catalyzes the cyclization of squalene into hopene, a key step in hopanoid (triterpenoid) metabolism. These enzymes are evolutionary related [1] proteins of about 70 to 85 Kd. As a signature pattern, a highly conserved region was selected which is rich in aromatic residues and which is located in the C-terminal section.

Consensus pattern: [DE]-G-S-W-x-G-x-W-[GA]-[LIVM]-x-[FY]-x-Y-[GA]

[1] Corey E.J., Matsuda S.P.T., Bartel B. Proc. Natl. Acad. Sci. U.S.A. 90:11628-11632(1993).

465. Prion protein signatures

Prion protein (PrP) [1,2,3] is a small glycoprotein found in high quantity in the brains of humans or animals infected with a number of degenerative neurological diseases such as Kuru, Creutzfeldt-Jacob disease (CJD), scrapie or bovine spongiform encephalopathy (BSE).

PrP is encoded in the host genome and expressed both in normal and infected cells. It has a tendency to aggregate yielding polymers called rods. Structurally, PrP is a protein consisting of a signal peptide, followed by an N-terminal domain that contains tandem repeats of a short motif (PHGGGWGQ in mammals, PHNPGY in chicken), itself followed by a highly conserved domain. It comes a C-terminal hydrophobic domain post-translationally removed when PrP is attached to the extracellular side of the cell membrane by a GPI-anchor. The structure of PrP is shown in the following schematic representation: +---+-----+---

*****-----* * * * +-----+ |Sig| Tandem repeats | C C S | | +---+-----+---
 -----|-----|-----+ +-----+ |GPIC: conserved cysteine involved in a

disulfide bond.¹⁸⁴; position of the patterns. As signature pattern for PrP, a perfectly conserved alanine- and glycine-rich region of 16 residues was selected as well as a region centered on the second cysteine involved in the disulfide bond.

- 5 Consensus pattern: A-G-A-A-A-G-A-V-V-G-G-L-G-G-Y-
Consensus pattern: E-x-[ED]-x-K-[LIVM](2)-x-[KR]-[LIVM](2)-x-[QE]-M-C-x(2)- Q-Y [C
is involved in a disulfide bond]

[1] Stahl N., Prusiner S.B. FASEB J. 5:2799-2807(1991).

- 10 [2] Brunori M., Chiara Silvestrini M., Pocchiari M. Trends Biochem. Sci. 13:309-313(1988).

[3] Prusiner S.B. Annu. Rev. Microbiol. 43:345-374(1989).

466. Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature and profile (pro
isomerase)

15 Cyclophilin [1] is the major high-affinity binding protein in vertebrates for the
immunosuppressive drug cyclosporin A (CSA). It exhibits a peptidyl- prolyl cis-trans
isomerase activity (EC 5.2.1.8) (PPIase or rotamase). PPIase is an enzyme that accelerates
protein folding by catalyzing the cis-transisomerization of proline imidic peptide bonds in
oligopeptides [2]. It is probable that CSA mediates some of its effects via an inhibitory action
20 on PPIase. Cyclophilin is a cytosolic protein which belongs to a family [3,4,5] that also
includes the following isozymes: - Cyclophilin B (or S-cyclophilin), a PPIase which is
retained in an endoplasmic reticulum compartment. - Cyclophilin C, a cytoplasmic PPIase. -
Mitochondrial matrix cyclophilin (cyp3). - A PPIase which seems specific for the folding of
25 rhodopsin and is an integral membrane protein anchored by a C-terminal transmembrane
region. This protein was first characterized in Drosophila (gene ninaA). - Bacterial
periplasmic PPIase (gene ppiA). - Bacterial cytosolic PPIase (gene ppiB). - Natural-killer cell
cyclophilin-related protein. This large protein (about 160 Kd) is a component of a putative
tumor-recognition complex involved in the function of NK cells. It contains a cyclophilin-
30 type PPIase domain. - Mammalian nucleoporin Nup358 [6], a nuclear pore complex protein
of 358 Kd that contains a C-terminal cyclophilin-type PPIase domain. - Yeast hypothetical
protein YJR032w. - Fission yeast hypothetical protein SpAC21E11.05c. - Caenorhabditis
elegans hypothetical protein T27D1.1. The sequences of the different forms of cyclophilin-

type PPIases are well conserved. As a signature pattern, a conserved region was selected in the central part of these enzymes.

Consensus pattern: [FY]-x(2)-[STCNLV]-x-F-H-[RH]-[LIVMN]-[LIVM]-x(2)-F- [LIVM]-x-

- 5 Q-[AG]-G- FKBP's, a family of proteins that bind the immunosuppressive drug FK506, are also PPIases, but their sequence is not at all related to that of cyclophilin.

[1] Stamnes M.A., Rutherford S.L., Zuker C.S. Trends Cell Biol. 2:272-276(1992).

[2] Fischer G., Schmid F.X. Biochemistry 29:2205-2212(1990).

- 10 [3] Trandinh C.C., Pao G.M., Saier M.H. Jr. FASEB J. 6:3410-3420(1992).

[4] Galat A. Eur. J. Biochem. 216:689-707(1993).

[5] Hacker J., Fischer G. Mol. Microbiol. 10:445-456(1993).

[6] Wu J., Matunis M.J., Kraemer D., Blobel G., Coutavas E. J. Biol. Chem. 270:14209-14213(1995).

15 467. Profilin signature

Profilin [1,2] is a small eukaryotic protein that binds to monomeric actin(G-actin) in a 1:1 ratio thus preventing the polymerization of actin into filaments (F-actin). It can also, in certain circumstance promotes actin polymerization. Profilin also binds to polyphosphoinositides such as PIP2. Overall sequence similarity among profilin from organisms which belong to different phyla (ranging from fungi to mammals) is low, but the N-terminal region is relatively well conserved. That region is thought to be involved in the binding to actin. The signature pattern for profilin is based on conserved residues at the N-terminal extremity .A protein structurally similar to profilin is present in the genome of variola and vaccinia viruses (gene A42R).

Consensus pattern: <x(0,1)-[STA]-x(0,1)-W-[DENQH]-x-[YI]-x-[DEQ]

- 30 [1] Haarer B.K., Brown S.S. Cell Motil. Cytoskeleton 17:71-74(1990).

[2] Sohn R.H., Goldschmidt-Clermont P. BioEssays 16:465-472(1994).

468. Protamine P1 signature

Protamines are small, highly basic proteins, that substitute for histones in sperm chromatin during the haploid phase of spermatogenesis. They pack sperm DNA into a highly condensed, stable and inactive complex. There are two different types of mammalian protamine, called P1 and P2. P1 has been found in all species studied, while P2 is sometimes absent. There seems to be a single type of avian protamine whose sequence is closely related to that of mammalian P1 [1]. As a signature for this family of proteins, a conserved region was selected at the N-terminal extremity of the sequence.

Consensus pattern: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S-

[1] Oliva R., Goren R., Dixon G.H. J. Biol. Chem. 264:17627-17630(1989).

469. Sperm histone P2 (protamine P2)

This protein also known as protamine P2 can substitute for histones in the chromatin of sperm. The alignment contains both the sequence of the mature P2 protein and its propeptide.

470. Proteasome A-type subunits signature

The proteasome (or macropain) (EC 3.4.99.46) [1 to 5, E1] is an eukaryotic and archaeobacterial multicatalytic proteinase complex that seems to be involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway. In eukaryotes the proteasome is composed of about 28 distinct subunits which form a highly ordered ring-shaped structure (20S ring) of about 700 Kd. Most proteasome subunits can be classified, on the basis of sequence similarities into two groups, A and B. Subunits that belong to the A-type group are proteins of from 210 to 290 amino acids that share a number of conserved sequence regions. Subunits that are known to belong to this family are listed below. - Vertebrate subunits C2 (nu), C3, C8, C9, iota and zeta. - Drosophila PROS-25, PROS-28.1, PROS-29 and PROS-35. - Yeast C1 (PRS1), C5 (PRS3), C7-alpha (Y8) (PRS2), Y7, Y13, PRE5, PRE6 and PUP2. - Arabidopsis thaliana subunits alpha and PSM30. - Thermoplasma acidophilum alpha-subunit. In this archaeobacteria the proteasome is composed of only two different subunits. As a

signature pattern for proteasome A-type subunits the best conserved region was selected, which is located in the N-terminal part of these proteins.

Consensus pattern: [FY]-x(4)-[STNV]-x-[FYW]-S-P-x-G-[RKH]-x(2)-Q-[LIVM]-[DE]-Y-
5 [SAD]-x(2)-[SAG]-. These proteins belong to family T1 in the classification of peptidases [6,E2].

[1] Rivett A.J. Biochem. J. 291:1-10(1993).

[2] Rivett A.J. Arch. Biochem. Biophys. 268:1-8(1989).

10 [3] Goldberg A.L., Rock K.L. Nature 357:375-379(1992).

[4] Wilk S. Enzyme Protein 47:187-188(1993).

[5] Hilt W., Wolf D.H. Trends Biochem. Sci. 21:96-102(1996).

[6] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

15 Proteasome B-type subunits signature

The proteasome (or macropain) (EC 3.4.99.46) [1 to 5,E1] is an eukaryotic and archaeobacterial multicatalytic proteinase complex that seems to be involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway. In eukaryotes the proteasome is composed of about 28 distinct subunits which form a highly ordered ring-shaped structure
20 (20S ring) of about 700 Kd. Most proteasome subunits can be classified, on the basis on sequence similarities into two groups, A and B. Subunits that belong to the B-type group are proteins of from 190 to 290 amino acids that share a number of conserved sequence regions. Subunits that are known to belong to this family are listed below. - Vertebrate subunits C5, beta, delta, epsilon, theta (C10-II), LMP2/RING12, C13 (LMP7/RING10), C7-I and MECL-
25 1. - Yeast PRE1, PRE2 (PRG1), PRE3, PRE4, PRS3, PUP1 and PUP3. - Drosophila L(3)73AI. - Fission yeast pts1. - Thermoplasma acidophilum beta-subunit. In this archaeobacteria the proteasome is composed of only two different subunits. As a signature pattern for proteasome B-type subunits the best conserved region was selected, which is located in the N-terminal part of these proteins.

30 Consensus pattern: [LIVMA]-[GSA]-[LIVMF]-x-[FYLVGAC]-x(2)-[GSACFY]-[LIVMSTAC](3)-[GAC]-[GSTACV]-[DES]-x(15)-[RK]-x(12,13)-G-x(2)-[GSTA]-D-. These proteins belong to family T1 in the classification of peptidases [6,E2].

- [1] Rivett A.J. *Biochem. J.* 291:1-10(1993).
 [2] Rivett A.J. *Arch. Biochem. Biophys.* 268:1-8(1989).
 [3] Goldberg A.L., Rock K.L. *Nature* 357:375-379(1992).
 5 [4] Wilk S. *Enzyme Protein* 47:187-188(1993).
 [5] Hilt W., Wolf D.H. *Trends Biochem. Sci.* 21:96-102(1996).
 [6] Rawlings N.D., Barrett A.J. *Meth. Enzymol.* 244:19-61(1994).

10 471. (pyr redox) Pyridine nucleotide-disulphide oxidoreductases class-I active site
 The pyridine nucleotide-disulphide oxidoreductases are FAD flavoproteins which contains a
 pair of redox-active cysteines involved in the transfer of reducing equivalents from the FAD
 cofactor to the substrate. On the basis of sequence and structural similarities [1] these
 enzymes can be classified into two categories. The first category groups together the
 15 following enzymes [2 to 6]: - Glutathione reductase (EC 1.6.4.2) (GR). - Higher eukaryotes
 thioredoxin reductase (EC 1.6.4.5). - Trypanothione reductase (EC 1.6.4.8). - Lipoamide
 dehydrogenase (EC 1.8.1.4), the E3 component of alpha-ketoacid dehydrogenase complexes.
 - Mercuric reductase (EC 1.16.1.1). The sequence around the two cysteines involved in the
 redox-active disulfide bond is conserved and can be used as a signature pattern.

20 Consensus pattern: G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P [The two C's form the active site
 disulfide bond]. In positions 6 and 7 of the pattern all known sequences have Asn-(Val/ Ile)
 with the exception of GR from plant chloroplasts and from cyanobacteria which have Ile-Arg
 [7].

- 25 [1] Kurlyan J., Krishna T.S.R., Wong L., Guenther B., Pahler A., Williams C.H. Jr., Model
 P. *Nature* 352:172-174(1991).
 [2] Rice D.W., Schulz G.E., Guest J.R. *J. Mol. Biol.* 174:483-496(1984).
 [3] Brown N.L. *Trends Biochem. Sci.* 10:400-402(1985).
 30 [4] Carothers D.J., Pons G., Patel M.S. *Arch. Biochem. Biophys.* 268:409-425(1989).
 [5] Walsh C.T., Bradley M., Nadeau K. *Trends Biochem. Sci.* 16:305-309(1991).
 [6] Gasdaska P.Y., Gasdaska J.R., Cochran S., Powis G. *FEBS Lett.* 373:5-9(1995).

[7] Creissen G., Edwards E.A., Enard C., Wellburn A., Mullineaux P. *Plant J.* 2:129-131(1991).

- 5 472. (pyridoxal deC) DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site (pyridoxal deC)

Three different enzymes - all pyridoxal-dependent decarboxylases - seem to share regions of sequence similarity [1,2,3,4], especially in the vicinity of the lysine residue which serves as the attachment site for the pyridoxal-phosphate (PLP) group. These enzymes are: - Glutamate decarboxylase (EC 4.1.1.15) (GAD). Catalyzes the decarboxylation of glutamate into the neurotransmitter GABA (4-aminobutanoate). - Histidine decarboxylase (EC 4.1.1.22) (HDC). Catalyzes the decarboxylation of histidine to histamine. There are two completely unrelated types of HDC: those that use PLP as a cofactor (found in Gram-negative bacteria and mammals), and those that contain a covalently bound pyruvoyl residue (found in Gram-positive bacteria). - Aromatic-L-amino-acid decarboxylase (EC 4.1.1.28) (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase. DDC catalyzes the decarboxylation of tryptophan to tryptamine. It also acts on 5-hydroxy- tryptophan and dihydroxyphenylalanine (L-dopa). - Tyrosine decarboxylase (EC 4.1.1.25) (TyrDC) which converts tyrosine into tyramine, a precursor of isoquinoline alkaloids and various amides. These enzymes are collectively known as group II decarboxylases [3,4].

Consensus pattern: S-[LIVMFYW]-x(5)-K-[LIVMFYWG](2)-x(3)-[LIVMFYW]-x-[CA]-x(2)-[LIVMFYWQ]-x(2)-[RK] [K is the pyridoxal-P attachment site]

- 25 [1] Jackson F.R. *J. Mol. Evol.* 31:325-329(1990).
 [2] Joseph D.R., Sullivan P., Wang Y.-M., Kozak C., Fenstermacher D.A., Behrendsen M.E., Zahnow C.A. *Proc. Natl. Acad. Sci. U.S.A.* 87:733-737(1990).
 [3] Sandmeier E., Hale T.I., Christen P. *Eur. J. Biochem.* 221:997-1002(1994).
 [4] Ishii S., Mizuguchi H., Nishino J., Hayashi H., Kagamiyama H. *J. Biochem.* 120:369-376(1996).
- 30

473. Regulator of chromosome condensation (RCC1) signatures (RCC1)

The regulator of chromosome condensation (RCC1) [1] is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein, to promote the loss of bound GDP and the uptake of fresh GTP, thus acting as a guanine-nucleotide dissociation stimulator (GDS) [2]. The interaction of RCC1 with ran probably plays an important role in the regulation of gene expression. RCC1, known as PRP20 or SRM1 in yeast, pim1 in fission yeast and BJI in *Drosophila*, is a protein that contains seven tandem repeats of a domain of about 50 to 60 amino acids. As shown in the following schematic representation, the repeats make up the major part of the length of the protein. Outside the repeat region, there is just a small N-terminal domain of about 40 to 50 residues and, in the *Drosophila* protein only, a C-terminal domain of about 130 residues.

In *Drosophila* two signature patterns for RCC1 were developed. The first is found in the N-terminal part of the second repeat; this is the most conserved part of RCC1. The second is derived from conserved positions in the C-terminal part of each repeat and detects up to five copies of the repeated domain. The RCC1-type of repeat is also found in the X-linked retinitis pigmentosa GTPase regulator [3].

Consensus pattern: G-x-N-D-x(2)-[AV]-L-G-R-x-T-

Consensus pattern: [LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM]-

[1] Dasso M. Trends Biochem. Sci. 18:96-101(1993).

[2] Boguski M.S., McCormick F. Nature 366:643-654(1993).

[3] Roepman R., Van Duijnhoven G., Rosenberg T., Pinckers A.J.L.G., Bleeker-

Wagemakers L.M., Bergen A.A.B., Post J., Beck A., Reinhardt R., Ropers H.-H., Cremers F., Berger W. Hum. Mol. Genet. 5:1035-1041(1996).

474. RNA 3'-terminal phosphate cyclase signature (RCT)

RNA 3'-terminal phosphate cyclase (EC [6.5.1.4](#)) [1,2] catalyzes the conversion of 3'-phosphate to a 2',3'-cyclic phosphodiester at the end of RNA. The biological role of this enzyme is unknown but it is likely to function in some aspects of cellular RNA processing. The reaction catalyzed by the enzyme occurs in three steps: 1) adenylation of the enzyme by

ATP; 2) the enzyme acts on RNA-3'terminal phosphate to produce RNA-3'terminal diphosphate adenylate; 3) Release of AMP and cyclisation by a non catalytic nucleophilic attack by the adjacent 2'hydroxyl on the phosphorus in the diester linkage. This enzyme, which has been characterized in human (where there seems to be at least three isozymes) and

5 Escherichia coli (gene *rtCA*), seems to be taxonomically widespread. It is found in insects, plants, fungi (gene *RTC1* in yeast) and in archaeobacteria. RNA cyclase is a protein of from 36 to 42 Kd. The best conserved region, which is used as a signature pattern, is a glycine-rich stretch of residues located in the central part of the sequence and which is reminiscent of various ATP, GTP or AMP glycine-rich loops. In this context, the conserved Arg (His in the

10 E.coli enzyme) could be the AMP-binding residue.

Consensus pattern: [RH]-G-x(2)-P-x-G(3)-x-[LIV]-

[1] Genschik P., Billy E., Swianiewicz M., Filipowicz W. EMBO J. 16:2955-2967(1997).

[2] Filipowicz W., Vincente O. Meth. Enzymol. 181:499-510(1990).

475. REV protein (anti-repression trans-activator protein)

476. Prokaryotic-type class I peptide chain release factors signature (RF-1)

Peptide chain release factors (RFs) are required for the termination of protein biosynthesis [1]. At present two classes of RFs can be distinguished. Class I RFs bind to ribosomes that have encountered a stop codon at their decoding site and induce release of the nascent

25 polypeptide. Class II RFs are GTP-binding proteins that interact with class I RFs and enhance class I RF activity. In prokaryotes there are two class I RFs that act in a codon specific manner[2]: RF-1 (gene *prfA*) mediates UAA and UAG-dependent termination while RF-2(gene *prfB*) mediates UAA and UGA-dependent termination. RF-1 and RF-2 are structurally and evolutionary related proteins which have been shown [3] to make up a family that also

30 contains the following proteins: - Fungal MRF1, a mitochondrial RF (m-RF) which recognizes the UAA and UAG codons. - Escherichia coli RF-H, a protein of unknown function. - Escherichia coli hypothetical protein *yaeJ* and a close *Pseudomonas putida*

homolog. A highly conserved region located in the central part of the 40 to 45 Kd RF-1/2 and m-RF and in the N-terminal of the 15 to 16Kd RF-H and yaeJ is used as a signature pattern.

Consensus pattern: [AR]-[STA]-x-G-x-G-G-Q-[HNGCS]-V-N-x(3)-[ST]-A-[IV]

- 5 Note that prokaryotic-type class I RFs display no significant sequence similarity to prokaryotic-type class II which belong to the family of GTP-binding elongation factors nor to eukaryotic class I or class II RFs.

10 [1] Tate W.P. , Poole E.S., Mannering S.M. Prog. Nucleic Acids. Res. Mol. Biol. 52:293-335(1996).

[2] Craigen W.J., Lee C.C., Caskey C.T. Mol. Microbiol. 4:861-865(1990).

[3] Pel H.J., Rep M., Grivell L.A. Nucleic Acids Res. 20:4423-4428(1992).

15 477. RIO1/ZK632.3/MJ0444 family signature

The following uncharacterized proteins are evolutionary related [1]: - Yeast protein RIO1. - *Caenorhabditis elegans* hypothetical protein ZK632.3. - *Methanococcus jannaschii* hypothetical protein MJ0444. - *Thermoplasma acidophilum* hypothetical protein if rpoA2 3'region. The eukaryotic members of this family are proteins of about 55 to 60 Kd, while the archeobacterial ones are half that size. The central part of these proteins is highly conserved. The best conserved region is used as a signature pattern.

Consensus pattern: [LIVM]-V-H-[GA]-D-L-S-E-[FY]-N-x-[LIVM]

25 [1] Bairoch A. Unpublished observations (1997).

30 478. (RIP) Shiga/ricin ribosomal inactivating toxins active site signature. A number of bacterial and plant toxins act by inhibiting protein synthesis in eukaryotic cells. The toxins of the Shiga and ricin family inactivate 60S ribosomal subunits by an N-glycosidic cleavage which releases a specific adenine base from the sugar-phosphate backbone of 28S rRNA [1,2,3]. The toxins which are known to function in this manner are: - Shiga toxin from *Shigella dysenteriae* [4]. This toxin is composed of one copy of an enzymatically active A

subunit and five copies of a B subunit responsible for binding the toxin complex to specific receptors on the target cell surface. - Shiga-like toxins (SLT) are a group of *Escherichia coli* toxins very similar in their structure and properties to Shiga toxin. The sequence of two types of these toxins, SLT-1 [5] and SLT-2 [6], is known. - Ricin, a potent toxin from castor bean seeds. Ricin consists of two glycosylated chains linked by a disulfide bond. The A chain is enzymatically active. The B chain is a lectin with a binding preference for galactosides. Both chains are encoded by a single polypeptidic precursor. Ricin is classified as a type-II ribosome-inactivating protein (RIP); other members of this family are agglutinin, also from castor bean, and abrin from the seeds of the bean *Abrus precatorius* [7]. - Single chain ribosome-inactivating proteins (type-I RIP) from plants. Examples of such proteins are: barley protein synthesis inhibitors I and II, mongolian snake-gourd trichosanthin, sponge gourd luffin-A and -B, garden four-o'clock MAP, common pokeberry PAP-S and soapwort saporin-6 [7]. All these toxins are structurally related. A conserved glutamic residue has been implicated [8] in the catalytic mechanism; it is located near a conserved arginine which also plays a role in catalysis [9]. The signature that has been developed for these proteins includes these catalytic residues.

Consensus pattern: [LIVMA]-x-[LIVMSTA](2)-x-E-[SAGV]-[STAL]-R-[FY]-[RKNQS]-x-[LIVM]-[EQS]-x(2)-[LIVMF] [E and R are active site residues]-

- [1] Endo Y., Tsurugi K., Takeda Y., Ogasawara T., Igarashi K. *Eur. J. Biochem.* 171:45-50(1988). [2] May M.J., Hartley M.R., Roberts L.M., Krieg P.A., Osborn R.W., Lord J.M. *EMBO J.* 8:301-308(1989). [3] Funatsu G., Islam M.R., Minami Y., Sung-Sil K., Kimura M. *Biochimie* 73:1157-1161(1991). [4] Strockbine N.A., Jackson M.P., Sung L.M., Holmes R.K., O'Brien A.D. *J. Bacteriol.* 170:1116-1122(1988). [5] Calderwood S.B., Auclair F., Donohue-Rolfe A., Keusch G.T., Mekalanos J.J. *Proc. Natl. Acad. Sci. U.S.A.* 84:4364-4368(1987). [6] Jackson M.P., Neill R.J., O'Brien A.D., Holmes R.K., Newland J.W. *FEMS Microbiol. Lett.* 44:109-114(1987). [7] Barbieri L., Battelli M.G., Stirpe F. *Biochim. Biophys. Acta* 1154:237-282(1993). [8] Hovde C.J., Calderwood S.B., Mekalanos J.J., Collier R.J. *Proc. Natl. Acad. Sci. U.S.A.* 85:2568-2572(1988). [9] Monzingo A.F., Collins E.J., Ernst S.R., Irvin J.D., Robertus J.D. *J. Mol. Biol.* 233:705-715(1993).

479. Bacterial RNA polymerase, alpha chain (RNA pol A bac)

Members of this family include alpha subunit from eubacteria and alpha subunits from chloroplasts. The alpha subunit of RNA polymerase consists of two independently folded domains, referred to as amino-terminal and carboxyl terminal domains. The amino terminal domain is involved in the interaction with the other subunits of the RNA polymerase. The carboxyl-terminal domain interacts with the DNA and activators. The amino acid sequence of the alpha subunit is conserved in prokaryotic and chloroplast RNA polymerases. There are three regions of particularly strong conservation, two in the amino-terminal and one in the carboxyl-Comment: terminal [3].

[1] Zhang G, Darst SA; Science 1998;281:262-266. [2] Jeon YH, Negishi T, Shirakawa M, Yamazaki T, Fujita N, Ishihama A, Kyogoku Y; Science 1995;270:1495-1497. [3] Ebright RH, Busby S; Curr Opin Genet Dev 1995;5:197-203. [4] Murakami K, Kimura M, Owens JT, Meares CF, Ishihama A; Proc Natl Acad Sci USA 1997;94:1709-1714.

480. RNA polymerase beta subunit (RNA pol B)

RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases). Each RNA polymerase complex contains two related members of this family, in each case they are the two largest subunits.

[1] Falkenburg D, Dworniczak B, Faust DM, Bautz EK; J Mol Biol 1987;195:929-937.

481. RNA polymerases H / 23 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaeobacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. Archaeobacterial subunit H (gene rpoH) [1,2] is a small protein of about 8.5 to 10 Kd, it is evolutionary related to the C-terminal part of a 23 Kd component shared by all three forms of eukaryotic RNA polymerases (gene RPB5 in yeast and POLR2E in mammals). As a signature pattern a conserved region was selected which is located at the N-terminal extremity of

subunit H; this region contains two histidines that could play a role in the binding of a metal ion.

Consensus pattern: H-[NEI]-[LIVM]-V-P-x-H-x(2)-[LIVM]-x(2)-[DE]

[1] Klenk H.-P., Palm P., Lottspeich F., Zillig W. Proc. Natl. Acad. Sci. U.S.A. 89:407-410(1992).

[2] Thiru A., Hodach M., Eloranta J.J., Kostourou V., Weinzierl R.O., Matthews S.; J. Mol. Biol. 287:753-760(1999).

482. RNA polymerases K / 14 to 18 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaeobacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. A component of 14 to 18 Kd shared by all three forms of eukaryotic RNA polymerases and which has been sequenced in budding yeast (gene RPB6 or RPO26), in fission yeast (gene rpb6 or rpo15), in human and in African swine fever virus [1] is evolutionary related [2] to archaeobacterial subunit K (gene rpoK). The archaeobacterial protein is colinear with the C-terminal part of the eukaryotic subunit.

Consensus pattern: [ST]-x-[FY]-E-x-[AT]-R-x-[LIVM]-[GSA]-x-R-[SA]-x-Q

[1] Lu Z., Kutish G.F., Sussman M.D., Rock D.L. Nucleic Acids Res. 21:2940-2940(1993).

[2] McKune K., Woychik N.A. J. Bacteriol. 176:4754-4756(1994).

483. RNA polymerases L / 13 to 16 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaeobacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. It

has been shown that small subunits of about 13 to 16 Kd found in all three types of eukaryotic polymerases are highly conserved. Subunits known to belong to this family are: - Budding yeast RPC19 subunit from RNA polymerases I and III [1]. - Budding yeast RPB11 subunit from RNA polymerase II [2]. - Mammalian RPB11 (gene POLR2K) from RNA polymerase II. - *Caenorhabditis elegans* hypothetical protein F58A4.9. - *Methanococcus jannaschii* RNA polymerase subunit L (gene rpoL). - *Sulfolobus acidocaldarius* RNA polymerase subunit L (gene rpoL) [3]. As a signature pattern a conserved region was selected which is located at the N-terminal extremity of these polymerase subunits; this region contains two cysteines that could play a role in the binding of a metal ion.

Consensus pattern: [DE](2)-H-[ST]-[LIVM]-[GAP]-N-x(11)-V-x-[FM]-x(2)-Y-x(3)-H-P

[1] Dequard-Chablat M., Riva M., Carles C., Sentenac A. J. Biol. Chem. 266:15300-15307(1991).

[2] Woychik N.A., McKune K., Lane W.S., Young R.A. Gene Expr. 3:77-82(1993).

[3] Langer D. EMBL/GenBank: X70805.

484. RNA polymerases N / 8 Kd subunits signature

In eukaryotes, there are three different forms of DNA-dependent RNA polymerases (EC 2.7.7.6) transcribing different sets of genes. Each class of RNA polymerase is an assemblage of ten to twelve different polypeptides. In archaeobacteria, there is generally a single form of RNA polymerase which also consist of an oligomeric assemblage of 10 to 13 polypeptides. Archaeobacterial subunit N (gene rpoN) [1] is a small protein of about 8 Kd, it is evolutionary related [2] to a 8.3 Kd component shared by all three forms of eukaryotic RNA polymerases (gene RPB10 in yeast and POLR2J in mammals) as well as to African swine fever virus protein CP80R [3]. As a signature pattern a conserved region was selected which is located at the N-terminal extremity of these polymerase subunits; this region contains two cysteines that could play a role in the binding of a metal ion.

Consensus pattern: [LIVMF](2)-P-[LIVM]-x-C-F-[ST]-C-G-

- [1] Langer D., Hain J., Thuriaux P., Zillig W. Proc. Natl. Acad. Sci. U.S.A. 92:5768-5772(1995).
- [2] McKune K., Woychik N.A. J. Bacteriol. 176:4754-4756(1994).
- [3] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).

485. Ribonuclease HII

- [1] Mian IS; Nucleic Acids Res 1997;25:3187-3189.

486. Ribonuclease PH signature

Prokaryotic ribonuclease PH (EC 2.7.7.56) (RNase PH) [1] is a phosphorolytic exoribonuclease that removes nucleotide residues following the -CCA terminus of tRNA and adds nucleotides to the ends of RNA molecules by using nucleoside diphosphates as substrates. RNase PH is a conserved protein of about 240 amino-acid residues. It is evolutionary related to *Caenorhabditis elegans* hypothetical protein B0564.1. As a signature pattern, the most highly conserved region was selected which is located in the central part of these proteins.

Consensus sequence: C-[DE]-[LIVM](2)-Q-[GTA]-D-G-[SG]-x(2)-[TA]-A

- [1] Kelly K.O., Deutscher M.P. J. Biol. Chem. 267:17153-17158(1992).

487. RanBP1 domain

- [1] Di Matteo G, Fuschi P, Zerfass K, Moretti S, Ricordy R, Cenciarelli C, Tripodi M, Jansen-Durr P, Lavia P; Cell Growth Differ 1995;6:1213-1224.

488. Rhodanese signatures

Rhodanese (thiosulfate sulfurtransferase) (EC 2.8.1.1) [1,2] is an enzyme which catalyzes the transfer of the sulfane atom of thiosulfate to cyanide, to form sulfite and thiocyanate. In vertebrates, rhodanese is a mitochondrial enzyme of about 300 amino-acid residues involved in forming iron-sulfur complexes and cyanide detoxification. A cysteine residue takes part in

the catalytic mechanism. Some bacterial proteins closely related to rhodanese are also thought to express a sulfotransferase activity. These are: - *Azotobacter vinelandii* rhdA. - *Escherichia coli* sseA [3]. - *Saccharopolyspora erythraea* cysA [4]. - *Synechococcus* strain PCC 7942 rhdA [5]. RhdA is a periplasmic protein probably involved in the transport of sulfur compounds. Two patterns for the rhodanese family were developed. They are based on highly conserved regions, one which is located in the N-terminal region, the other at the C-terminal extremity of the enzyme.

Consensus pattern: [FY]-x(3)-H-[LIV]-P-G-A-x(2)-[LIVF]

Consensus pattern: [FY]-[DEAP]-G-[SA]-W-x-E-[FYW]

[1] Westley J. Meth. Enzymol. 77:285-291(1981).

[2] Weiland K.L., Dooley T.P. Biochem. J. 275:227-231(1991).

[3] Rudd K.E. Unpublished observations (1993).

[4] Donadio S., Shafiee A., Hutchinson C.R. J. Bacteriol. 172:350-360(1990).

[5] Laudenbach D.E., Ehrhardt D., Green L., Grossman A.R. J. Bacteriol. 173:2751-2760(1991).

489. Ribonuclease III family signature

Prokaryotic ribonuclease III (EC 3.1.26.3) (gene *rnc*) [1] is an enzyme that digests double-stranded RNA. It is involved in the processing of ribosomal RNA precursors and of some mRNAs. RNase III is evolutionary related [2] to the following proteins: - Fission yeast *pac1*, a ribonuclease that probably inhibits mating and meiosis by degrading a specific mRNA required for sexual development. - Yeast ribonuclease III (gene *RNT1*), a dsRNA-specific nuclease that cleaves eukaryotic preribosomal RNA at various sites. - *Caenorhabditis elegans* hypothetical protein F26E4.13. - *Paramecium bursaria* chloroella virus 1 protein A464R. - *Synechocystis* strain PCC 6803 hypothetical protein slr0346. - Fission yeast hypothetical protein SpAC8A4.08c, a protein with a N-terminal helicase domain and a C-terminal RNase III domain. - *Caenorhabditis elegans* hypothetical protein K12H4.8, a protein with the same structure as SpAC8A4.08c. These proteins share regions of sequence similarity; one of which is a highly conserved stretch of 9 residues which has been developed as a signature pattern.

Consensus pattern: [DEQ]-[RQ]-[LM]-E-[FYW]-[LV]-G-D-[SAR]-

[1] Nashimoto H., Uchida H. Mol. Gen. Genet. 201:25-29(1985).

[2] Mian I.S. Nucleic Acids Res. 25:3187-3195(1997).

5

490. Rieske iron-sulfur protein signatures

Ubiquinol-cytochrome c reductase (EC 1.10.2.2) (also known as the bc1 complex or complex III) is one of the electron transport chains of mitochondria and of some aerobic prokaryotes; it catalyzes the oxidoreduction of ubiquinol and cytochrome c. In the chloroplast of plants and in cyanobacteria plastoquinone-plastocyanin reductase (EC 1.10.99.1) (also known as the b6f complex) is functionally similar and catalyzes the oxidoreduction of plastoquinol and cytochrome f. One of the components of these electron transfer systems is an iron-sulfur protein with a 2Fe-2S cluster, which is called the Rieske protein [1,2]. The Rieske protein contains approximately 190 amino acid residues. The iron-sulfur cluster is complexed to the protein through cysteine and histidine residues. Two perfectly conserved regions in Rieske proteins contain all the residues that bind the iron-sulfur cluster. Both regions contain two cysteines and a histidine. The first cysteine and the histidine are 2Fe-2S ligands while the remaining cysteines form a disulfide bond [3]. Two conserved regions were selected as signature patterns.

10

15

20

Consensus pattern: C-[TK]-H-L-G-C-[LIVST] [The first C and the H are 2Fe-2S ligands]
[The second C is involved in a disulfide bond]

Consensus pattern: C-P-C-H-x-[GSA] [The first C and the H are 2Fe-2S ligands] [The second C is involved in a disulfide bond]

25

[1] Gatti F.L., Meinhardt S.W., Ohnishi T., Tzagoloff A. J. Mol. Biol. 205:421-435(1989).

[2] Kallas T., Spiller S., Malkin R. Proc. Natl. Acad. Sci. U.S.A. 85:5794-5798(1988).

[3] Iwata S., Saynovits M., Link T.A., Michel H. Structure 4:567-579(1996).

30

491. Ribosomal protein L1 signature

Ribosomal protein L1 is the largest protein from the large ribosomal subunit. In *Escherichia coli*, L1 is known to bind to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1, 2], groups: - Eubacterial L1. - Algal and plant chloroplast L1. - Cyanelle L1. - Archaeobacterial L1. - Vertebrate L10A. - Yeast SSM1. As a signature pattern, the best conserved region was selected located in the central section of these proteins. It is located at the end of an alpha helix thought to be involved in RNA-binding.

Consensus pattern: [IM]-x(2)-[LIVA]-x(2,3)-[LIVM]-G-x(2)-[LMS]-[GSNH]-[PTKR]-[KRAV]-G-x-[LIMF]-P-[DENSTKQ]

[1] Nikonov S.V., Nevskaya N., Eliseikina I.A., Fomenkova N.P., Nikulin A., Ossina N., Garber M., Jonsson B.-H., Briand C., Al-Karadaghi S., Svensson L.A., Aevansson A., Liljas A. EMBO J. 15:1350-1359(1996).

[2] Olvera J., Wool I.G. 2,3.CO:2-"Biochem. Biophys. Res. Commun. 220:954-957(1996).

492. Ribosomal protein L10 signature

Ribosomal protein L10 is one of the proteins from the large ribosomal subunit. L10 is a protein of 162 to 185 amino-acid residues which has only been found so far in eubacteria. A conserved region located in the N-terminal section of these proteins was used as a signature pattern.

Consensus pattern: [DEH]-x(2)-[GS]-[LIVMF]-[STN]-[VA]-x-[DEQK]-[LIVMA]-x(2)-[LIM]-R

493. Ribosomal protein L10e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Vertebrate L10 (QM) [1]. - Plant L10. - *Caenorhabditis elegans* L10 (F10B5.1). - Yeast L10 (QSR1). - *Methanococcus jannaschii* MJ0543. These proteins have 174 to 232 amino-acid residues. A conserved region located in the central section was selected as a signature pattern.

Consensus pattern: R-x-A-[FYW]-G-K-[PA]-x-G-x(2)-A-R-V

[1] Chan Y.-L., Díaz J.-J., Denoroy L., Madjar J.-J., Wool I.G. 2.3.CO;2-"Biochem. Biophys. Res. Commun. 255:952-956(1996).

494. Ribosomal protein L11 signature

Ribosomal protein L11 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L11 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups:

- Eubacterial L11.
- Plant chloroplast L11 (nuclear-encoded).
- Read algal chloroplast L11.
- Cyanelle L11.
- Archaeobacterial L11.
- Mammalian L12.
- Plants L12.
- Yeast L12 (YL15).

L11 is a protein of 140 to 165 amino-acid residues. A conserved region located in the C-terminal section of these proteins was selected as a signature pattern. In Escherichia coli, the C-terminal half of L11 has been shown [3] to be in an extended and loosely folded conformation and is likely to be buried within the ribosomal structure.

Consensus pattern: [RKN]-x-[LIVM]-x-G-[ST]-x(2)-[SNQ]-[LIVM]-G-x(2)-[LIVM]-x(0,1)-[DENG]

- [1] Pucciarelli G., Remacha M., Ballesta J.P.G.; Nucleic Acids Res. 18:4409-4416(1990).
 [2] Otaka E., Hashimoto T., Mizuta K., Suzuki K.; Protein Seq. Data Anal. 5:301-313(1993).
 [3] Choli T. Biochem. Int. 19:1323-1338(1989).

495. Ribosomal protein L7/L12 C-terminal domain

[1] Leijonmarck M, Liljas A; J Mol Biol 1987;195:555-579.

496. Ribosomal protein L13 signature

Ribosomal protein L13 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L13 is known to be one of the early assembly proteins of the 50S ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L13.

- Plant chloroplast L13 (nuclear-encoded). - Red algal chloroplast L13.

- Archaeobacterial L13. - Mammalian L13a (Tum P198). - Yeast Rp22 and Rp23.

L11 is a protein of 140 to 250 amino-acid residues. As a signature pattern, a conserved region was selected located in the C-terminal section of these proteins.

Consensus pattern: [LIVM]-[KRV]-[GK]-M-[LIV]-[PS]-x(4,5)-[GS]-[NQEKRA]-x(5)-[LIVM]-x-[AIV]-[LFY]-x-[GDN]

[1] Chan Y.-L., Olvera J., Glueck A., Wool I.G. J. Biol. Chem. 269:5589-5594(1994).

497. Ribosomal protein L13e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Vertebrate L13 (was previously known as Breast Basic Conserved protein 1 (BBC1)). - *Drosophila* L13. - Plant L13. - Yeast probable L13 (YM9375.11c).

These proteins have 199 to 218 amino-acid residues. As a signature pattern, a stretch of about 16 residues in the first third of these proteins selected.

-Consensus pattern: [KR]-Y-x(2)-K-[LIVM]-R-[STA]-G-[KR]-G-F-[ST]-L-x-E

[1] Olvera J., Wool I.G. Biochem. Biophys. Res. Commun. 201:102-107(1994).

498. Ribosomal protein L14 signature

5 Ribosomal protein L14 is one of the proteins from the large ribosomal subunit.

In eubacteria, L14 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L14. - Algal and plant chloroplast L14. - Cyanelle L14. - Archaeobacterial L14. - Yeast L17A. - Mammalian L23.

10 - *Caenorhabditis elegans* L23 (B0336.10). - Higher eukaryotes mitochondrial L14. - Yeast mitochondrial Yml38 (gene MRPL38).

L14 is a protein of 119 to 137 amino-acid residues. As a signature pattern, a conserved region located in the C-terminal half of these proteins was selected.

15 -Consensus pattern: [GA]-[LIV](3)-x(9,10)-[DNS]-G-x(4)-[FY]-x(2)-[NT]-x(2)-V-[LIV]

[1] Otake E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

499. Ribosomal protein L15 signature

Ribosomal protein L15 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L15 is known to bind the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1],

25 groups: - Eubacterial L15. - Plant chloroplast L15 (nuclear-encoded).

- Archaeobacterial L15. - Vertebrate L27a. - *Tetrahymena thermophila* L29.

- Fungi L27a (L29, CRP-I, CYH2).

L15 is a protein of 144 to 154 amino-acid residues. As a signature pattern, a conserved region was selected in the C-terminal section of these proteins.

30

-Consensus pattern: K-[LIVM](2)-[GASL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-[LIVMFCA]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G

[1] Otake E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

500. Ribosomal protein L15e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian L15. - Insect L15. - Plant L15. - Yeast YL10 (L13) (Rp15r).
- Thermoplasma acidophilum L15.

These proteins have about 200 amino acid residues. As a signature pattern, a conserved region was selected located in the central section.

-Consensus pattern: [DE]-[KR]-A-R-x-L-G-[FY]-x-[SAP]-x(2)-G-[LIVMFY](4)-R-x-R-[IV]-x-R-G

[1] Zwickl P., Lupas A., Baumeister W.

Biochem. Biophys. Res. Commun. 209:684-688(1995).

501. Ribosomal protein L17 signature

Ribosomal protein L17 is one of the proteins from the large ribosomal subunit.

L17 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L17.

- Yeast mitochondrial YmL8 (gene MRPL8).

Eubacterial L17 is a protein of 120 to 130 amino-acid residues. Yeast YmL8 is twice larger (238 residues), the sequence of its N-terminal half is colinear with that of eubacterial L17. As a signature pattern, a conserved region in the N-terminal section was selected.

-Consensus pattern: I-x-[ST]-[GT]-x(2)-[KR]-x-K-x(6)-[DE]-x-[LIMV]-[LIVMT]-T-x-[STAG]-[KR]

502. Ribosomal protein L18e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

445

- Vertebrate L18 (known as L14 in *Xenopus*) [1]. - Plant L18.

- Yeast L18 (Rp28). - *Halobacterium marismortui* H129.

- *Sulfolobus acidocaldarius* H129e.

These proteins have 115 to 187 amino-acid residues., A stretch of about 13 residues in the first third of these proteins has been selected as a signature pattern.

-Consensus pattern: [KRE]-x-L-x(2)-[PS]-[KR]-x(2)-[RH]-[PSA]-x-[LIVM]-[NS]-[LIVM]-x-[RK]-[LIVM]

[1] Puder M., Barnard G.F., Staniunas R.J., Steele G.D. Jr., Chen L.B.

Biochim. Biophys. Acta 1216:134-136(1993).

503. Ribosomal L18p family

It has been shown that the amino terminal 93 amino acids of Swiss:P09895 are necessary and sufficient to bind 5S rRNA in vitro. The carboxyl-terminal half of the protein, comprising amino acids 151-296, serves to localize the protein to the nucleolus [1].

Number of members: 26

[1]

Medline: 96212235

Distinct domains in ribosomal protein L5 mediate 5 S rRNA binding and nucleolar localization.

Michael WM, Dreyfuss G;

J Biol Chem 1996;271:11571-11574.

504. Ribosomal protein L19 signature

Ribosomal protein L19 is one of the proteins from the large ribosomal subunit. In *Escherichia coli*, L19 is known to be located at the 30S-50S ribosomal subunit interface and may play a role in the structure and function of the aminoacyl-tRNA binding site. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L19.

- Red algal chloroplast L19. - *Cyanelle* L19.

L19 is a protein of 120 to 130 amino-acid residues.,

A conserved region in the C-terminal section has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x-[KRGTI]-x-[GSAI]-[KRODA]-[VG]-[RSN]-X(0,1)-[KR]-
[SA]-[KY]-[KLI]-[LYS]-Y-[LIM]-R

5

505. Ribosomal protein L19e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- 10 - Mammalian ribosomal protein L19 [1]. - Drosophila ribosomal protein L19 [2].
- Slime mold (*D. discoideum*) vegetative specific protein V14 [3].

- Yeast ribosomal protein L19 (YL14). - Archeobacterial ribosomal protein L19E.
These proteins have 148 to 203 amino-acid residues.

- 15 A stretch of about 20 residues in the N-terminal part of these
proteins has been selected as a signature pattern.

-Consensus pattern: Q-[KR]-R-[LIVM]-x-[SA]-x(4)-[CV]-G-x(3)-[IV]-[WK]-[LIVF]-
[DN]-P

[1] Chan Y.-L., Lin A., McNally J., Peleg D., Meyuhos O., Wool I.G.

J. Biol. Chem. 262:1111-1115(1987).[2] Hart K., Klein T., Wilcox M.

- 20 Mech. Dev. 43:101-110(1993).[3] Singleton C.K., Manning S.S., Ken R.
Nucleic Acids Res. 17:9679-9692(1989).

506. Ribosomal protein L1e signature (Ribosomal_L4)

- 25 A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped
on the basis of sequence similarities. One of these families consists [1,2,3,
4] of: Vertebrate L1 (L4). - Drosophila L1. - Plant L1. - Yeast L2 (Rp2).

- Fission yeast L2. - Halobacterium marismortui HmaL4 (HL6).

- Methanococcus jannaschii MJ0177.

- 30 These proteins have 246 (archaeobacteria) to 427 (human) amino acids. A conserved region
in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: N-x(3)-[KRM]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-
x(7)-[RK]-[GS]-H

[1] Rafti F., Gargiulo G., Manzi A., Malva C., Graziani F.

Nucleic Acids Res. 17:456-456(1989).[2] Presutti C., Villa T., Bozzoni I.

Nucleic Acids Res. 21:3900-3900(1993).

[3] Bagni C., Mariottini P., Annesi F., Amaldi F.

5 Biochim. Biophys. Acta 1216:475-478(1993).

[3] Arndt E., Kroemer W., Hatakeyama T. J. Biol. Chem. 265:3034-3039(1990).

507. Ribosomal protein L2 signature

10 Ribosomal protein L2 is one of the proteins from the large ribosomal subunit.

In Escherichia coli, L2 is known to bind to the 23S rRNA and to have peptidyltransferase activity. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial L2.

- Algal and plant chloroplast L2. - Cyanelle L2. - Archaeobacterial L2.

15 - Plant L2. - Slime mold L2. - Marchantia polymorpha mitochondrial L2.

- Paramecium tetraurelia mitochondrial L2. - Fission yeast K5, K37 and KD4.

- Yeast YL6. - Vertebrate L8.

The best conserved region located in the C-terminal section of these proteins has been selected as

20 a signature pattern.

-Consensus pattern: P-x(2)-R-G-[STAIV](2)-x-N-[APK]-x-[DE]

[1] Marty I., Meyer Y.

Nucleic Acids Res. 20:1517-1522(1992).

[2] Otaka E., Hashimoto T., Mizuta K., Suzuki K.

25 Protein Seq. Data Anal. 5:301-313(1993).

508. Ribosomal protein L20 signature

Ribosomal protein L20 is one of the proteins from the large ribosomal subunit.

30 In Escherichia coli, L20 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L20. - Algal and plant chloroplast L20.

- Cyanelle L20.

L20 is a protein of about 120 amino-acid residues. A conserved region located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: K-x(3)-[KRC]-x-[LIVM]-W-[IV]-[STNALV]-R-[LIVM]-[NS]-x(3)-[RKHS]

- 5 [1] Otake E., Hashimoto T., Mizuta K., Suzuki K.
Protein Seq. Data Anal. 5:301-313(1993).

509. Ribosomal protein L21e signature

- 10 A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L21 [1]. - *Entamoeba histolytica* L21 [2].
- *Caenorhabditis elegans* L21 (C14B9.7). - Yeast L21E (URP1) [3].
- *Halobacterium marismortui* HL31 [4].

- 15 These proteins have 160 (eukaryotes) or 95 (archeobacteria) amino-acid residues. A conserved region in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: G-[DE]-x-V-x(10)-[GV]-x(2)-[FYH]-x(2)-[FY]-x-G-x-T-G

- 20 [1] Devi K.R.G., Chan Y.-L., Wool I.G.
Biochem. Biophys. Res. Commun. 162:364-370(1989).
[2] Petter R., Rozenblatt S., Nuchamowitz Y., Mirelman D.
Mol. Biochem. Parasitol. 56:329-333(1992).
[3] Jank B., Waldherr M., Schweyen R.J. Curr. Genet. 23:15-18(1993).
[4] Hatakeyama T., Kimura M. Eur. J. Biochem. 172:703-711(1988).

25

510. Ribosomal protein L21 signature

Ribosomal protein L21 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L21 is known to bind to the 23S rRNA in the presence of

- 30 L20. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups:

- *Marchantia polymorpha* chloroplast L21. - *Cyanella* L21.
- Spinach chloroplast L21 (nuclear-encoded).

Eubacterial L21 is a protein of about 100 amino-acid residues, the mature form of the spinach chloroplast L21 has 200 residues. A conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [IVT]-x(3)-[KR]-x(3)-[KRO]-K-x(6)-G-[HF]-R-[RO]-x(2)-[ST]

5

511. Ribosomal protein L22 signature

Ribosomal protein L22 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L22 is known to bind 23S rRNA. It belongs to a family of

10 ribosomal proteins which, on the basis of sequence similarities [1,2,3],

groups: - Eubacterial L22.

- Algal and plant chloroplast L22 (in legumes L22 is encoded in the nucleus instead of the chloroplast). - Cyanelle L22. - Archaeobacterial L22.

- Mammalian L17. - Plant L17. - Yeast YL17.

15 A conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [RKQN]-x(4)-[RH]-[GAS]-x-G-[KRQS]-x(9)-[HDN]-[LIVM]-x-[LIVMS]-x-[LIVM]

[1] Gantt J.S., Baldauf S.L., Calie P.J., Weeden N.F., Palmer J.D.

20 EMBO J. 10:3073-3078(1991).[2] Madsen L.H., Kreiberg J.D., Gausing K.

Curr. Genet. 19:417-422(1991).

[3] Otaka E., Hashimoto T., Mizuta K., Suzuki K.

Protein Seq. Data Anal. 5:301-313(1993).

25

512. Ribosomal protein L23 signature

Ribosomal protein L23 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L23 is known to bind a specific region on the 23S rRNA;

in yeast, the corresponding protein binds to a homologous site on the 26S rRNA

30 [1]. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [2,3,4], groups: - Eubacterial L23.

- Algal and plant chloroplast L23. - Archaeobacterial L23. - Mammalian L23A.

- *Caenorhabditis elegans* L23A (F55D10.2). - Fungi L25.

- Yeast mitochondrial YmL41 (gene MRPL41 or MRP20).

A small conserved region in the C-terminal section of these proteins, which is probably involved in rRNA-binding has been selected as a signature pattern [2].

-Consensus pattern: [RK](2)-[AM]-[IVFYT]-[IV]-[RKT]-L-[STANEOK]-x(7)-[LIVMFT]

[1] El Baradi T.T.A.L., Raue H.A., van de Regt C.H.F., Verbree E.C.,
Planta R.J. EMBO J. 4:210-2107(1985).

[2] Raue H.A., Otake E., Suzuki K. J. Mol. Evol. 28:418-426(1989).

[3] Fearon K., Mason T.L. J. Biol. Chem. 267:5162-5170(1992).

[4] Otake E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

513. Ribosomal protein L24 signature

Ribosomal protein L24 is one of the proteins from the large ribosomal subunit. L24 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L24.

- Plant chloroplast L24 (nuclear-encoded). - Red algal L24. - Vertebrate L26.

- Yeast L26 (YL33). - Archaeobacterial HmaL24 (HL15).

- A probable ribosomal protein from Sulfolobus acidocaldarius [1].

In their mature form, these proteins have 103 to 150 amino-acid residues.

A conserved stretch of 20 residues in their N-terminal section has been selected as a signature pattern.

-Consensus pattern: [GDEN]-D-x-V-x-[IV]-[LIVMA]-x-G-x(2)-[KRA]-[GNQ]-x(2,3)-

[GA]-x-[IV]

[1] Ouzounis C., Kyripides N., Sander C.

Nucleic Acids Res. 23:565-570(1995).

514. Ribosomal protein L24e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists [1] of:

- Mammalian ribosomal protein L24.

451

- Yeast ribosomal protein L30A/B (Rp29) (YL21).
- *Kluyveromyces lactis* ribosomal protein L30.
- *Arabidopsis thaliana* ribosomal protein L24 homolog.
- *Haloarcula marismortui* ribosomal protein HL21/HL22.

5 - *Methanococcus jannaschii* MJ1201.

These proteins have 60 to 160 amino-acid residues. The most conserved region, which is located in the N-terminal region of these proteins has been selected as a signature pattern.

-Consensus pattern: [FY]-x-[GSH]-x(2)-[IV]-x-P-G-x-G-x(2)-[FYV]-x-[KRHE]-x-D

[1] Chan Y.-L., Olvera J., Wool I.G.

10 Biochem. Biophys. Res. Commun. 202:1176-1180(1994).

515. Ribosomal protein L27 signature

Ribosomal protein L27 is one of the proteins from the large ribosomal subunit.

15 L27 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial L27.

- Plant chloroplast L27 (nuclear-encoded). - Algal chloroplast L27.
- Yeast mitochondrial YmL2 (gene MRPL2 or MRP7).

The schematic relationship between these groups of proteins is shown below.

20 Eub. L27 NxxxxxxxxAlgal L27 Nxxxxxxxx

Plant L27 ttttNxxxxxxxxxxxx

Yeast MRP7 tttNxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

***t': transit peptide.

'N': N-terminal of mature protein. 't': position of the pattern.

25 -Consensus pattern: G-x-[LIVM](2)-x-R-Q-R-G-x(5)-G

[1] Elhag G.A., Bourque D.P. Biochemistry 31:6856-6864(1992).

[2] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

30

516. Ribosomal L28 family

The ribosomal 28 family includes L28 proteins from bacteria and chloroplasts. The L24 protein from yeast Swiss:P36525

452

also contains a region of similarity to prokaryotic L28 proteins. L24 from yeast is also found in the large ribosomal subunit

Number of members: 24

5

517. Ribosomal protein L29 signature

Ribosomal protein L29 is one of the proteins from the large ribosomal subunit.

L29 belongs to a family of ribosomal proteins which, on the basis of sequence

10 similarities [1], groups: - Eubacterial L29. - Red algal L29.

- Archaeobacterial L29. - Mammalian L35 - *Caenorhabditis elegans* L35 (ZK652.4).

- Yeast L35.

L29 is a protein of 63 to 138 amino-acid residues.

A conserved region located in the central section of L29 has been selected as a
signature pattern.

-Consensus pattern: [KNQS]-[PSTL]-x(2)-[LIMFA]-[KRGSA]-x-[LIVYSTA]-[KR]-
[KRHQS]-[DESTANRL]-[LIV]-A-[KRCQVT]-[LIVMA]

[1] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

20

518. Ribosomal protein L3 signature

Ribosomal protein L3 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L3 is known to bind to the 23S rRNA and may participate

25 in the formation of the peptidyltransferase center of the ribosome. It belongs

to a family of ribosomal proteins which, on the basis of sequence

similarities [1,2,3,4], groups: - Eubacterial L3. - Red algal L3. - *Cyanella* L3.

- Archaeobacterial *Halobacterium marismortui* HmaL3 (HL1).

- Yeast L3 (also known as trichodermin resistance protein) (gene TCM1).

30 - *Arabidopsis thaliana* L3 (genes ARP1 and ARP2). - Mammalian L3 (L4).

- Mammalian mitochondrial L3. - Yeast mitochondrial YmL9 (gene MRPL9).

A conserved region located in the central section of these proteins has been selected
as a signature pattern.

-Consensus pattern: [FL]-x(6)-[DN]-x(2)-[AGS]-x-[ST]-x-G-[KRH]-G-x(2)-G-x(3)-R

[1] Arndt E., Kroemer W., Hatakeyama T. J. Biol. Chem. 265:3034-3039(1990).

[2] Graack H.-R., Grohmann L., Kitakawa M., Schaefer K.L., Kruft V.

Eur. J. Biochem. 206:373-380(1992).

5 [3] Herwig S., Kruft V., Wittmann-Liebold B.

Eur. J. Biochem. 207:877-885(1992).

[4] Otake E., Hashimoto T., Mizuta K., Suzuki K.

Protein Seq. Data Anal. 5:301-313(1993).

10 519. Ribosomal protein L30 signature

Ribosomal protein L30 is one of the proteins from the large ribosomal subunit.

L30 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L30. - Archaeobacterial L30.

15 - Drosophila L7. - Slime mold L7. - Mammalian L7. - Fungi L7 (YL8).

- Yeast mitochondrial L33.

L30 from eubacteria are small proteins of about 60 residues, those from archaeobacteria are proteins of about 150 residues. Eukaryotic L7 are proteins of about 250 to 270 residues. The schematic relationship between the three

20 groups of proteins is shown below. Eub. L30 NxxxxxxxxxxC

Arc. L30 NxxxxxxxxxxxxxxxxxxxxxxxxxxC

Euk. L7 NxxC

* * * * * ! * ! : position of the pattern.

The signature pattern for this family of ribosomal proteins spans the

25 N-terminal half of the region common to all these proteins.

-Consensus pattern: [IVT]-[LIVM]-x(2)-[LF]-x-[LI]-x-[KRHOEG]-x(2)-[STNQH]-x-

[IVT]-x(10)-[LMS]-[LIV]-x(2)-[LIVA]-x(2)-[LMFY]-[IVT]

[1] Mizuta K., Hashimoto T., Otake E.

Nucleic Acids Res. 20:1011-1016(1992).

30 520. Ribosomal protein L31 signature

Ribosomal protein L31 is one of the proteins from the large ribosomal subunit.

L31 is a protein of 66 to 97 amino-acid residues which has only been found so far in eubacteria and in some algal chloroplasts.

A conserved region located in the central section of these proteins has been selected as a signature pattern.

5 -Consensus pattern: H-P-F-[FY]-[TI]-x(9)-G-R-[AIV]-x-[KRO]

521. Ribosomal protein L31e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L31 [1]. - *Chlamydomonas reinhardtii* L31. - Yeast L34.
- *Halobacterium marismortui* HL30 [2].

These proteins have 87 to 128 amino-acid residues.

A conserved region, located in the central section has been selected as a signature pattern.

15 -Consensus pattern: V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AKH]-x-W-x-[KR]-G

[1] Tanaka T., Kuwano Y., Kuzumaki T., Ishikawa K., Ogata K.

Eur. J. Biochem. 162:45-48(1987).[2] Bergmann U., Arndt E.

Biochim. Biophys. Acta 1050:56-60(1990).

522. Ribosomal protein L33 signature

Ribosomal protein L33 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L33 has been shown to be on the surface of 50S subunit.

L33 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial L33.

- Algal and plant chloroplast L33. - Cyanelle L33.

L33 is a small protein of 49 to 66 amino-acid residues. A conserved region located in the central section of L33 has been selected as a signature pattern.

-Consensus pattern: Y-x-[ST]-x-[KR]-[NS]-x(4)-[PATQ]-x(1,2)-[LIVM]-[EA]-x(2)-

30 K-[FY]-[CSD]

[1] Kruff V., Kapp U., Wittmann-Liebold B. Biochimie 73:855-860(1991).

[2] Sharp P.M. Gene 139:129-130(1994).

[3] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

523. Ribosomal protein L34 signature

Ribosomal protein L34 is one of the proteins from the large subunit of the prokaryotic ribosome. It is a small basic protein of 44 to 51 amino-acid residues [1]. L34 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, groups: - Eubacterial L34.

- Red algal chloroplast L34. - Cyanelle L34.

A conserved region that corresponds to the N-terminal half of L34 has been selected as a signature pattern.

-Consensus pattern: K-[RG]-T-[FYWL]-[EQS]-x(5)-[KRHS]-x(4,5)-G-F-x(2)-R

[1] Old I.G., Margarita D., Saint Girons I.

Nucleic Acids Res. 20:6097-6097(1992).

524. Ribosomal protein L34e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L34. - Mosquito L31 [1]. - Plant L34 [2].

- Yeast putative ribosomal protein YIL052c. - Methanococcus jannaschii MJ0655.

These proteins have 89 to 129 amino-acid residues.

A conserved region located in the N-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: Y-x-[ST]-x-S-[NY]-x(5)-[KR]-T-P-G

[1] Lan Q., Niu L.L., Fallon A.M.

Biochim. Biophys. Acta 1218:460-462(1994).

[2] Gao J., Kim S.R., Chung Y.Y., Lee J.M., An G.

Plant Mol. Biol. 25:761-770(1994).

525. Ribosomal protein L35Ae signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

456

- Vertebrate L35A. - *Caenorhabditis elegans* L35A (F10E7.7).

- Yeast L37A/L37B (Rp47). - *Pyrococcus woesei* L35A homolog [1].

These proteins have 87 to 110 amino-acid residues.

A highly conserved stretch of 22 residues in the C-terminal part of

5 these proteins has been selected as a signature pattern.

-Consensus pattern: G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P

[1] Ouzounis C., Kyriades N., Sander C.

Nucleic Acids Res. 23:565-570(1995).

10

526. Ribosomal protein L36 signature

Ribosomal protein L36 is the smallest protein from the large subunit of the prokaryotic ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial L36. - Algal and plant chloroplast L36. - Cyanelle L36. L36 is a small basic and cysteine-rich protein of 37 amino-acid residues. As a signature pattern, a conserved region that corresponds to positions 11 to 36 in L36 and includes three conserved cysteine residues has been developed.

Consensus pattern: C-x(2)-C-x(2)-[LIVM]-x-R-x(3)-[LIVMN]-x-[LIVM]-x-C-x(3,4)-[KR]-H-x-Q-x-Q-

[1] Otake E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

20

527. Ribosomal protein L36e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian L36 [1].

- *Drosophila* L36 (M(1)1B). - *Caenorhabditis elegans* L36 (F37C12.4).

- *Candida albicans* L39. - Yeast YL39.

These proteins have 99 to 104 amino acids.

A conserved region in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR]

[1] Chan Y.-L., Paz V., Olvera J., Wool I.G.

Biochem. Biophys. Res. Commun. 192:849-853(1993).

30

528. Ribosomal protein L39e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L39 [1]. - Plants L39. - Yeast L46 [2]. - Archeobacterial L39e [3].

These proteins are very basic. About 50 residues long, they are the smallest proteins of eukaryotic-type ribosomes. A conserved region in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [KRA]-T-x(3)-[LIVM]-[KROF]-x-[NHS]-x(3)-R-[NHY]-W-R-R

[1] Lin A., McNally J., Wool I.G. J. Biol. Chem. 259:487-490(1984).

[2] Leer R.J., van Raamsdonk-Duin M.M.C., Kraakman P., Mager W.H., Planta R.J. Nucleic Acids Res. 13:701-709(1985).

[3] Ramirez C., Louie K.A., Matheson A.T. FEBS Lett. 250:416-418(1989).

529. Ribosomal L40e family

Bovine L40 has been identified as a secondary RNA binding protein [1]. L40 is fused to a ubiquitin protein [2].

Number of members: 27

[1]

Medline: 88203200

RNA binding proteins of the large subunit of bovine mitochondrial ribosomes.

Piatyszek MA, Denslow ND, O'Brien TW;

Nucleic Acids Res 1988;16:2565-2583.

[2]Medline: 96011832

The carboxyl extensions of two rat ubiquitin fusion proteins are ribosomal proteins S27a and L40.

Chan YL, Suzuki K, Wool IG;

Biochem Biophys Res Commun 1995;215:682-690.

530. (Ribosomal L44) Ribosomal protein L44e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L44 [1]. - *Trypanosoma brucei* L44.
- *Caenorhabditis elegans* L44 (C09H10.2). - Fungal L44 (L41).
- *Halobacterium marismortui* LA [2].

These proteins have 92 to 105 amino-acid residues.

A conserved region located in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: K-x-[TV]-K-K-x(2)-L-[KR]-x(2)-C

[1] Gallagher M.J., Chan Y.-L., Lin A., Wool I.G. DNA 7:269-273(1988).

[2] Bergmann U., Wittmann-Liebold B.

Biochim. Biophys. Acta 1173:195-200(1993)

531. Ribosomal protein L5 signature

Ribosomal protein L5 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L5 is known to be involved in binding 5S RNA to the large ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3,4], groups: - Eubacterial L5.

- Algal chloroplast L5. - Cyanelle L5. - Archaeobacterial L5. - Mammalian L11.
- *Tetrahymena thermophila* L21. - Slime mold L5 (V18). - Yeast L16 (39A).
- Plants mitochondrial L5.

L5 is a protein of about 180 amino-acid residues.

A conserved region, located in the first third of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x(2)-[LIVM]-[STAVC]-[GE]-[QV]-x(2)-[LIVMA]-x-[STC]-x-[STAG]-[KRH]-x-[STA]

[1] Hatakeyama T., Hatakeyama T. Biochim. Biophys. Acta 1039:343-347(1990).

[2] Rosendahl G., Andreassen P.H., Kristiansen K. Gene 98:161-167(1991).

[3] Yang D., Gunther L., Matheson A.T., Auer J., Spicker G., Boeck A.

Biochimie 73:679-682(1991).

[4] Otaka E., Hashimoto T., Mizuta K., Suzuki K.

Protein Seq. Data Anal. 5:301-313(1993).

532. ribosomal L5P family C-terminus

5 This region is found associated with Ribosomal_L5.

Number of members: 60

533. Ribosomal protein L6 signatures

10 Ribosomal protein L6 is one of the proteins from the large ribosomal subunit. In Escherichia coli, L6 is known to bind directly to the 23S rRNA and is located at the aminoacyl-tRNA binding site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3,4], groups: - Eubacterial L6.

- 15 - Algal chloroplast L6.
- Cyanelle L6.
- Archaeobacterial L6.
- Marchantia polymorpha mitochondrial L6.
- Yeast mitochondrial YmL6 (gene MRPL6).
- 20 - Mammalian L9.
- Drosophila L9.
- Plants L9.
- Yeast L9 (YL11).

25 While all the above proteins are evolutionary related it is very difficult to derive a pattern that will find them all. Two patterns were therefore created, the first to detect eubacterial, cyanelle and mitochondrial L6, the second to detect archaeobacterial L6 as well as eukaryotic L9.

-Consensus pattern: [PS]-[DENS]-x-Y-K-[GA]-K-G-[LIVM]

-Consensus pattern: Q-x(3)-[LIVM]-x(2)-[KR]-x(2)-R-x-F-x-D-G-[LIVM]-Y-[LIVM]-x(2)-
30 [KR]

[1] Suzuki K., Olvera J., Wool I.G. Gene 93:297-300(1990).

[2] Schwank S., Harrer R., Schueller H.-J., Schweizer E. Curr. Genet. 24:136-140(1993).

[3] Golden B.L., Ramakrishnan V., White S.W. EMBO J. 12:4901-4908(1993).

[4] Otaka E., Hashimoto T., Mizuta K., Suzuki K. Protein Seq. Data Anal. 5:301-313(1993).

5 534. Ribosomal protein L6e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian ribosomal protein L6 (L6 was previously known as TAX-responsive enhancer element binding protein 107).

10 - *Caenorhabditis elegans* ribosomal protein L6 (R151.3).

- Yeast ribosomal protein YL16A/YL16B.

- *Mesembryanthemum crystallinum* ribosomal protein YL16-like.

These proteins have 175 (yeast) to 287 (mammalian) amino acids. A highly conserved region in the central part of these proteins has been selected as a signature pattern.

15 - Consensus pattern: N-x(2)-P-L-R-R-x(4)-[FY]-V-I-A-T-S-x-K

535. Ribosomal protein L7Ae signature

20 A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate L7A (SURF3) [1]. - Plant L7A. - Yeast L7A (YL5) (Rp6).

- Yeast protein NHP2 [2]. - Yeast hypothetical protein YEL026w.

- *Bacillus subtilis* hypothetical protein ylxQ. - *Halobacterium marismortui* Hs6.

25 - *Methanococcus jannaschii* MJ1203.

These proteins have 100 to 265 amino-acid residues.

A conserved region located in the central section has been selected as a signature pattern.

- Consensus pattern: [CA]-x(4)-[IV]-P-[FY]-x(2)-[LIVM]-x-[GSQ]-[KRO]-x(2)-L-G

[1] Colombo P., Yon J., Garson K., Fried M.

30 Proc. Natl. Acad. Sci. U.S.A. 89:6358-6362(1992).

[2] Kolodrubetz D., Burgum A. Yeast 7:79-90(1991).

536. Ribosomal protein L9 signature

Ribosomal protein L9 is one of the proteins from the large ribosomal subunit.

In *Escherichia coli*, L9 is known to bind directly to the 23S rRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities

[1,2], groups: - Eubacterial L9. - Cyanobacterial L9.

- Plant chloroplast L9 (nuclear-encoded). - Red algal chloroplast L9.

A conserved region, located in the N-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: G-x(2)-[GN]-x(4)-V-x(2)-G-[FY]-x(2)-N-[FY]-L-x(5)-[GA]-

x(3)-[STN]

[1] Hoffman D.W., Davies C., Gerchman S.E., Kycia J.H., Porter S.J.,

White S.W., Ramakrishnan V. EMBO J. 13:205-212(1994).

[2] Otaka E., Hashimoto T., Mizuta K., Suzuki K.

Protein Seq. Data Anal. 5:301-313(1993).

537. Ribosomal protein S10 signature

Ribosomal protein S10 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S10 is known to be involved in binding tRNA to the

ribosomes. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S10.

- Algal chloroplast S10. - Cyanelle S10. - Archaeobacterial S10.

- Marchantia polymorpha and Prototheca wickerhamii mitochondrial S10.

- Arabidopsis thaliana mitochondrial S10 (nuclear encoded). - Vertebrate S20.

- Plant S20. - Yeast URP2.

S10 is a protein of about 100 amino-acid residues.

A conserved region located in the center of these proteins has been selected as a signature pattern.

-Consensus pattern: [AV]-x(3)-[GDNSR]-[LIVMSTA]-x(3)-G-P-[LIVM]-x-[LIVM]-P-T

[1] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

538. Ribosomal protein S11 signature

Ribosomal protein S11 [1] plays an essential role in selecting the correct tRNA in protein biosynthesis. It is located on the large lobe of the small ribosomal subunit. S11 belongs to a family of ribosomal proteins which, on the

basis of sequence similarities, groups [2]: - Eubacterial S11.

- Algal and plant chloroplast S11. - Cyanelle S11. - Archaeobacterial S11.

- Marchantia polymorpha and Prototheca wickerhamii mitochondrial S11.

- Acanthamoeba castellanii mitochondrial S11. - Neurospora crassa S14 (crp-2).

- Yeast S14 (RP59 or CRY1).

- Mammalian, Drosophila, Trypanosoma, and plant S14.

- Caenorhabditis elegans S14 (F37C12.9).

One of the best conserved regions in these proteins was selected as a signature pattern.

-Consensus pattern: [LIVMF]-x-[GSTAC]-[LIVMF]-x(2)-[GSTAL]-x(0,1)-[GSN]-

[LIVMF]-x-[LIVM]-x(4)-[DEN]-x-T-P-x-[PA]-[STCH]-[DN]

[1] Kimura M., Kimura J., Hatakeyama T. FEBS Lett. 240:15-20(1988).

[2] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

539. Ribosomal protein S12 signature

Ribosomal protein S12 is one of the proteins from the small ribosomal subunit.

In Escherichia coli, S12 is known to be involved in the translation initiation step. It is a very basic protein of 120 to 150 amino-acid residues. S12

belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S12. - Archaeobacterial S12.

- Algal and plant chloroplast S12. - Cyanelle S12.

- Protozoa and plant mitochondrial S12. - Yeast S28.

- Drosophila mitochondrial protein tko (Technical KnockOut). - Mammalian S23.

The best conserved regions in these proteins, located in the center of each sequence have been selected as a signature pattern.

-Consensus pattern: [RK]-x-P-N-S-[AR]-x-R

[1] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

540. Ribosomal protein S12e signature

- 5 A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Vertebrate S12 [1].
- Trypanosoma brucei S12 [2]. - Caenorhabditis elegans S12 (F54E7.2).
- Drosophila S12. - Yeast S12.

These proteins have 130 to 150 amino acids.

- 10 A conserved region in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: A-L-[KRQP]-x-V-L-x(2)-[SA]-x(3)-[DN]-G-L

[1] Lin A., Chan Y.-L., Jones R., Wool I.G.

J. Biol. Chem. 262:14343-14351(1987).[2] Marchal C., Ismaili N., Pays E.

15 Mol. Biochem. Parasitol. 57:331-334(1993).

541. Ribosomal protein S13 signature

Ribosomal protein S13 is one of the proteins from the small ribosomal subunit.

- 20 In Escherichia coli, S13 is known to be involved in binding fMet-tRNA and, hence, in the initiation of translation. It is a basic protein of 115 to 177 amino-acid residues and belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S13.

- Plant chloroplast S13 (nuclear encoded). - Red algal chloroplast S13.

- 25 - Cyanelle S13. - Archaeobacterial S13. - Plant mitochondrial S13.

- Mammalian and plant S18.

The best conserved regions in these proteins, located in their C-terminal part have been selected as a signature pattern.

-Consensus pattern: [KRQS]-G-x-R-H-x(2)-[GSNH]-x(2)-[LIVMC]-R-G-Q

- 30 [1] Chan Y.-L., Paz V., Wool I.G.

Biochem. Biophys. Res. Commun. 178:1212-1218(1991).

[2] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

542. Ribosomal protein S14p/S29e (Ribosomal protein S14 signature)

Ribosomal protein S14 is one of the proteins from the small ribosomal subunit. In *Escherichia coli*, S14 is known to be required for the assembly of 30S particles and may also be responsible for determining the conformation of 16S rRNA at the A site. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups:

- Eubacterial S14.
- Algal and plant chloroplast S14.
- Cyanelle S14.
- Archaeobacterial *Methanococcus vannielii* S14.
- Plant mitochondrial S14.
- Yeast mitochondrial MRP2.
- Mammalian S29.
- Yeast YS29A/B.

S14 is a protein of 53 to 115 amino-acid residues. Our signature pattern is based on the few conserved positions located in the center of these proteins.

Consensus pattern: [RP]-x(0,1)-C-x(11,12)-[LIVMF]-x-[LIVMF]-[SC]-[RG]-x(3)-[RN]

[1] Chan Y.-L., Suzuki K., Olvera J., Wool I.G. *Nucleic Acids Res.* 21:649-655(1993).

[2] Otake E., Hashimoto T., Mizuta K. *Protein Seq. Data Anal.* 5:285-300(1993).

543. Ribosomal protein S15 signature

Ribosomal protein S15 is one of the proteins from the small ribosomal subunit. In *Escherichia coli*, this protein binds to 16S ribosomal RNA and functions at early steps in ribosome assembly. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups:

- Eubacterial S15.
- Archaeobacterial *Halobacterium marismortui* HmaS15 (HS11).
- Plant chloroplast S15. - Yeast mitochondrial S28. - Mammalian S13.
- *Brugia pahangi* and *Wuchereria bancrofti* S13 (S15). - Yeast S13 (YS15).

S15 is a protein of 80 to 250 amino-acid residues.

A conserved region located in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x(2)-H-[LIVMFY]-x(5)-D-x(2)-[SAGN]-x(3)-[LF]-x(9)-[LIVM]-x(2)-[FY]

5 [1] Dang H., Ellis S.R.

Nucleic Acids Res. 18:6895-6901(1990).

[2] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

10

544. Ribosomal protein S16 signature

Ribosomal protein S16 is one of the proteins from the small ribosomal subunit. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

15

- Eubacterial S16.
- Algal and plant chloroplast S16.
- Cyanelle S16.
- Neurospora crassa mitochondrial S24 (cyt-21).

20

S16 is a protein of about 100 amino-acid residues. A conserved region located in the N-terminal extremity of these proteins has been selected as a signature pattern.

Consensus pattern: [LIVMT]-x-[LIVM]-[KR]-L-[STAK]-R-x-G-[AKR]

[1] Otake E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

25

545. Ribosomal protein S17 signature

Ribosomal protein S17 is one of the proteins from the small ribosomal subunit.

30

In *Escherichia coli*, S17 is known to bind specifically to the 5' end of 16S ribosomal RNA and is thought to be involved in the recognition of termination codons. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups:

- Eubacterial S17.
- Plant chloroplast S17 (nuclear encoded).
- Red algal chloroplast S17.

- Cyanelle S17. - Archaeobacterial S17. - Mammalian and plant cytoplasmic S11.
- Yeast S18a and S18b (RP41; YS12).

The best conserved regions located in the C-terminal sections of these proteins have been selected as a signature pattern.

-Consensus pattern: G-D-x-[LIV]-x-[LIVA]-x-[QEK]-x-[RK]-P-[LIV]-S

[1] Gantt J.S., Thompson M.D. J. Biol. Chem. 265:2763-2767(1990).

[2] Herfurth E., Hirano H., Wittmann-Liebold B.

Biol. Chem. Hoppe-Seyler 372:955-961(1991).

[3] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

546. Ribosomal protein S17c signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrates S17 [1]. - Drosophila S17 [2]. - Neurospora crassa S17 (crp-3).

- Yeast S17a (RP51A) and S17b (RP51B) [3]. - Methanococcus jannaschii MJ0245.

These proteins have from 63 (in archeobacteria) to 130 to 146 amino acids and are highly conserved. A region in the central part of these proteins has been selected as a signature.

-Consensus pattern: A-x-I-x-[ST]-K-x-L-R-N-[KR]-I-A-G-[FY]-x-T-H

[1] Chen I.-T., Roufa D.J. Gene 70:107-116(1988).

[2] Maki C., Rhoads D.D., Stewart M.J., van Slyke B., Denell R.E.,

Roufa D.J. Gene 79:289-298(1989).[3] Abovich N., Rosbash M.

Mol. Cell. Biol. 4:1871-1879(1984).

547. Ribosomal protein S18 signature

Ribosomal protein S18 is one of the proteins from the small ribosomal subunit. In

Escherichia coli, S18 has been involved in aminoacyl-tRNA binding[1]. It appears to be situated at the tRNA A-site of the ribosome. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities[2], groups: - Eubacterial S18. - Algal and plant chloroplast S18. - Cyanelle S18. As a signature pattern, a conserved region in the central

section of the protein has been selected. This region contains two basic residues which may be involved in RNA-binding.-

Consensus pattern: [IV]-[DY]-Y-x(2)-[LIVMT]-x(2)-[LIVM]-x(2)-[FYT]-[LIVM]-[ST]-[DERP]-x-[GY]-K-[LIVM]-x(3)-R-[LIVMAS]-

- 5 [1] McDougall J., Choli T., Kruft V., Kapp U., Wittmann-Liebold B. FEBS Lett. 245:253-260(1989).[2] Otake E., Hashimoto T., Mizuta K. Protein Seq. Data Anal. 5:285-300(1993).

548. Ribosomal protein S19 signature

10 Ribosomal protein S19 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S19 is known to form a complex with S13 that binds strongly to 16S ribosomal RNA. S19 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S19.

- Algal and plant chloroplast S19. - Cyanelle S19. - Archaeobacterial S19.

15 - Plant mitochondrial S19. - Eukaryotic S15 ('rig' protein).

S19 is a protein of 88 to 144 amino-acid residues. Our signature pattern is based on the few conserved positions located in the C-terminal section of these proteins.

-Consensus pattern: [STDNQ]-G-[KMQ]-x(6)-[LIVM]-x(4)-[LIVM]-[GSD]-x(2)-[LF]-[GAS]-[DE]-F-x(2)-[ST]

20 [1] Kitagawa M., Takasawa S., Kikuchi N., Itoh T., Teraoka H., Yamamoto H., Okamoto H. FEBS Lett. 283:210-214(1991).

[2] Otake E., Hashimoto T., Mizuta K.
Protein Seq. Data Anal. 5:285-300(1993).

25

549. Ribosomal protein S19e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities [1,2]. One of these families consists

30 of: - Mammalian S19. - *Drosophila* S19.

- *Ascaris lumbricoides* S19g (ALEP-1) and S19s. - Yeast YS16 (RP55A and RP55B).

- *Aspergillus* S16. - *Halobacterium marismortui* HS12.

These proteins have 143 to 155 amino acids.

A well conserved stretch of 20 residues in the C-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: P-x(6)-[SAN]-x(2)-[LIVMA]-x-R-x-[ALIV]-[LV]-Q-x-L-[EQ]

[1] Etter A., Aboutanos M., Tobler H., Mueller F.

5 Proc. Natl. Acad. Sci. U.S.A. 88:1593-1596(1991).

[2] Suzuki K., Olvera J., Wool I.G. Biochimie 72:299-302(1990).

550. Ribosomal protein S2 signatures

10 Ribosomal protein S2 is one of the proteins from the small ribosomal subunit.

S2 belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S2. - Algal and plant chloroplast S2.

- Cyanelle S2. - Archaeobacterial S2.

- Higher eukaryotes P40 (previously thought to be a laminin receptor).

15 - Yeast NAB1. - Plant mitochondrial S2. - Yeast mitochondrial MRP4.

S2 is a protein of 235 to 394 amino-acid residues.

Two conserved regions have been selected as signature patterns. One is located in the N-terminal section and the other in the central section.

-Consensus pattern: [LIVMFA]-x(2)-[LIVMFYC](2)-x-[STAC]-[GSTANQEKR]-[STALV]-
20 [HY]-[LIVMF]-G

-Consensus pattern: P-x(2)-[LIVMF](2)-[LIVMS]-x-[GDN]-x(3)-[DENL]-x(3)-[LIVM]-
x-E-x(4)-[GNQKRH]-[LIVM]-[AP]

[1] Davis S.C., Tzagoloff A., Ellis S.R.

J. Biol. Chem. 267:5508-5514(1992).

25 [2] Tohgo A., Takasawa S., Munakata H., Yonekura H., Hayashi N., Okamoto H.

FEBS Lett. 340:133-138(1994).

551. Ribosomal protein S21 signature

30 Ribosomal protein S21 is one of the proteins from the small ribosomal subunit. So far

S21 has only been found in eubacteria. It is a protein of 55 to 70 amino-acid residues. A conserved region in the N-terminal section of the protein has been selected as a signature pattern.

Consensus pattern: [DE]-x-A-[LIY]-[KR]-R-F-K-[KR]-x(3)-[KR]

552. Ribosomal protein S21e signature

- 5 A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian S21 [1].
- *Caenorhabditis elegans* S21 (F37C12.11). - Rice S21 [2].
- Yeast S21 (Ys25) [3]. - Fission yeast S28 [4].

These proteins have 82 to 87 amino acids.

- 10 A perfectly conserved nonapeptide in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: L-Y-V-P-R-K-C-S-[SA]

[1] Bhat K.S., Morrison S.G. Nucleic Acids Res. 21:2939-2939(1993).

[2] Nishi R., Hashimoto H., Uchimiya H., Kato A.

- 15 Biochim. Biophys. Acta 1216:113-114(1993).[3] Suzuki K., Otaka E.

Nucleic Acids Res. 16:6223-6223(1988).[4] Itoh T., Okata E., Matsui K.A.

Biochemistry 24:7418-7423(1985).

- 20 553. Ribosomal protein S24e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Vertebrate S24 [1]. - Yeast Rp50. - *Mucor racemosus* S24 [2].

- *Halobacterium marismortui* HS15 [3]. - *Methanococcus jannaschii* MJ0394.

- 25 These proteins have 101 to 148 amino acids.

A well conserved stretch in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: [FYA]-G-x(2)-[KR]-[STA]-x-G-[FY]-[GA]-x-[LIVM]-Y-[DN]-[SDN]

- 30 [1] Brown S.J., Jewell A., Maki C.G., Roufa D.J. Gene 91:293-296(1990).

[2] Sosa L., Fonzi W.A., Sypherd P.S.

Nucleic Acids Res. 17:9319-9331(1989).[3] Kimura J., Arndt E., Kimura M.

FEBS Lett. 224:65-70(1987).

554. Ribosomal protein S26e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of: - Mammalian S26 [1].
 - Octopus S26 [2]. - Drosophila S26 (DS31) [3]. - Plant cytoplasmic S26.
 - Fungi S26 [4].

These proteins have 114 to 127 amino acids.

A conserved octapeptide in the central part of these proteins has been selected as a signature pattern.

-Consensus pattern: [YH]-C-V-S-C-A-I-H

[1] Kuwano Y., Nakanishi O., Nabeshima Y., Tanaka T., Ogata K.
 J. Biochem. 97:983-992(1985).[2] Zinov'eva R.D., Tomarev S.I.
 Dokl. Akad. Nauk SSSR 304:464-469(1989).

[3] Itoh N., Ohta K., Ohta M., Kawasaki T., Yamashina I.
 Nucleic Acids Res. 17:2121-2121(1989).[4] Wu M., Tan H.
 Gene 150:401-402(1994).

555. Ribosomal protein S28e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S28 [1]. - Plant S28 [2]. - Fungi S33 [3].
 - Methanococcus jannaschii MJ1202.

These proteins have from 64 to 78 amino acids.

A highly conserved nonapeptide from the C-terminal extremity of these proteins has been selected as a signature pattern.

-Consensus pattern: E-[ST]-E-R-E-A-R-x-L

[1] Chan Y.-L., Olvera J., Wool I.G.

Biochem. Biophys. Res. Commun. 179:314-318(1991).

[2] Hwang I., Goodman H.M. Plant Physiol. 102:1357-1358(1993).

[3] Hoekstra R., Ferreira P.M., Bootsman T.C., Mager W.H., Planta R.J.
 Yeast 8:949-959(1992).

556. Ribosomal protein S3Ae signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S3A (was originally known as v-fos transformation effector protein).
- *Caenorhabditis elegans* S3A (F56F3.5).
- Plant cytoplasmic S3A (CYC07) [1].
- Yeast Rp10 (PLC1 and PLC2).
- Fission yeast Rp10 (SpAC13G6.02c).
- *Methanococcus jannaschii* MJ0980.

These proteins have from 220 to 250 amino acids.

A conserved stretch in their N-terminal section was selected as a signature pattern.

-Consensus pattern: [LIV]-x-[GH]-R-[IV]-x-E-x-[SC]-L-x-D-L

[1] Liu J.H., Reid D.M.

Plant Physiol. 109:338-338(1995).

557. Ribosomal protein S3 signature

Ribosomal protein S3 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial S3.
- Algal and plant chloroplast S3.
- Cyanelle S3.
- Archaeobacterial S3.
- Plant mitochondrial S3.
- Vertebrate S3.
- Insect S3.
- *Caenorhabditis elegans* S3 (C23G10.3).
- Yeast S3 (Rp13).

S3 is a protein of 209 to 559 amino-acid residues.

A conserved region located in the C-terminal section has been selected as a signature pattern.

-Consensus pattern: [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[DENQ]-x(7)-[LMT]-x(2)-G-x(2)-G

[1] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

558. Ribosomal protein S4 signature

Ribosomal protein S4 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S4 is known to bind directly to 16S ribosomal RNA.

Mutations in S4 have been shown to increase translational error frequencies.

It belongs to a family of ribosomal proteins which, on the basis of sequence

similarities [1,2], groups: - Eubacterial S4. - Algal and plant chloroplast S4.

- Cyanelle S4. - Archaeobacterial S4. - Mammalian S9. - Yeast YS11 (SUP45).

- *Marchantia polymorpha* mitochondrial S4. - *Dictyostelium discoideum* rp1024.

- Yeast protein NAM9 [3]. NAM9 has been characterized as a suppressor for

ochre mutations in mitochondrial DNA. It could be a ribosomal protein that

acts as a suppressor by decreasing translation accuracy.

S4 is a protein of 171 to 205 amino-acid residues (except for NAM9 which is

much larger). The signature pattern for this protein is based on a conserved

region located in the central section of these proteins.

-Consensus pattern: [LIVM]-[DE]-x-R-[LI]-x(3)-[LIVMC]-[VMFYHQ]-[KRT]-x(3)-

[STAGCVF]-x-[ST]-x(3)-[SAI]-[KR]-x-[LIVMF](2)

[1] Mizuta K., Hashimoto T., Suzuki K.I., Otaka E.

Nucleic Acids Res. 19:2603-2608(1991).

[2] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

[3] Boguta M., Dmochowska A., Borsuk P., Wrobel K., Gargouri A., Lazowska J.,

Slonimski P., Szczesniak B., Kruszezewska A.

Mol. Cell. Biol. 12:402-412(1992).

559. Ribosomal protein S4e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S4 [1]. Two highly similar isoforms of this protein exist: one coded by a gene on chromosome Y, and the other on chromosome X.

- Plant cytoplasmic S4 [2] - Yeast S7 (YS6). - Archeobacterial S4e.

These proteins have 233 to 264 amino acids.

A highly conserved stretch of 15 residues in their N-terminal section has been selected as a signature pattern. Four positions in this region are positively

charged residues.

-Consensus pattern: H-x-K-R-[LIVMF]-[SANK]-x-P-x(2)-[WY]-x-[LIVM]-x-[KRP]

[1] Fisher E.M., Beer-Romero P., Brown L.G., Ridley A., McNeil J.A.,

Lawrence J.B., Willard H.F., Bieber F.R., Page D.C.

5 Cell 63:1205-1218(1990).

[2] Braun H.P., Emmermann M., Mentzel H., Schmitz U.K.

Biochim. Biophys. Acta 1218:435-438(1994).

10 560. Ribosomal protein S5 signature

Ribosomal protein S5 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S5 is known to be important in the assembly and function of the 30S ribosomal subunit. Mutations in S5 have been shown to increase translational error frequencies. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S5.

- Cyanelle S5. - Red algal chloroplast S5. - Archaeobacterial S5.

- Mammalian S2 (LLrep3). - *Caenorhabditis elegans* S2 (C49H3.11).

- *Drosophila* S2. - Plant S2. - Yeast S4 (SUP44). - Fungi mitochondrial S5.

S5 is a protein of 166 to 254 amino-acid residues. The signature pattern for

20 this protein is based on a conserved region, rich in glycine residues, and located in the N-terminal section of these proteins.

-Consensus pattern: G-[KRQ]-x(3)-[FY]-x-[ACV]-x(2)-[LIVMA]-[LIVM]-[AG]-[DN]-
x(2)-G-x-[LIVM]-G-x-[SAG]-x(5,6)-[DEQ]-[LIVMA]-x(2)-A-
[LIVMF]

25 [1] All-Robyn J.A., Brown N., Otaka E., Liebman S.W.

Mol. Cell. Biol. 10:6544-6553(1990).[2] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

30 561. Ribosomal protein S6 signature

Ribosomal protein S6 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S6 is known to bind together with S18 to 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of

sequence similarities, groups: - Eubacterial S6. - Red algal chloroplast S6.
- Cyanelle S6.

S6 is a protein of 95 to 208 amino-acid residues. The signature pattern for this protein is based on a conserved region located in the N-terminal section of these proteins.

-Consensus pattern: G-x-[KRC]-[DENQRH]-L-[SA]-Y-x-I-[KRNSA]

562. Ribosomal protein S6e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian S6 [1]. - Drosophila S6 [2]. - Plant S6 [3]. - Yeast S10 (YS4).
- Halobacterium marismortui HS13 [4]. - Methanococcus jannaschii MJ1260.

S6 is the major substrate of protein kinases in eukaryotic ribosomes [5]; it may have an important role in controlling cell growth and proliferation through the selective translation of particular classes of mRNA.

These proteins have 135 to 249 amino acids.

A conserved stretch of 12 residues in the N-terminal part of these proteins has been selected as a signature pattern.

-Consensus pattern: [LIVM]-[STAMR]-G-G-x-D-x(2)-G-x-P-M

[1] Franco R., Rosenfeld M.G. J. Biol. Chem. 265:4321-4325(1990).

[2] Watson K.L., Konrad K.D., Woods D.F., Bryant P.J.
Proc. Natl. Acad. Sci. U.S.A. 89:11302-11306(1992).

[3] Hansen G., Estruch J.J., Spena A.
Nucleic Acids Res. 20:5230-5230(1992).

[4] Kimura M., Arndt E., Hatakeyama T., Hatakeyama T., Kimura J.
Can. J. Microbiol. 35:195-199(1989).

[5] Bandi H.R., Ferrari S., Krieg J., Meyer H.E., Thomas G.
J. Biol. Chem. 268:4530-4533(1993).

563. Ribosomal protein S7 signature

Ribosomal protein S7 is one of the proteins from the small ribosomal subunit.

In *Escherichia coli*, S7 is known to bind directly to part of the 3' end of 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2,3], groups: - Eubacterial S7.

- Algal and plant chloroplast S7. - Cyanelle S7. - Archaeobacterial S7.

- Plant mitochondrial S7. - Mammalian S5. - Plant S5.

- *Caenorhabditis elegans* S5 (T05E11.1).

The best conserved region located in the N-terminal section of these proteins has been selected as a signature pattern.

- Consensus pattern: [DENSK]-x-[LIVMDET]-x(3)-[LIVMFTA](2)-x(6)-G-K-[KR]-x(5)-

[LIVMF]-[LIVMFC]-x(2)-[STAC]

[1] Klussmann S., Franke P., Bergmann U., Kostka S., Wittmann-Liebold B.

Biol. Chem. Hoppe-Seyler 374:305-312(1993).

[2] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

[3] Ignatovich O., Cooper M., Kulesza H.M., Beggs J.D.

Nucleic Acids Res. 23:4616-4619(1995).

564. Ribosomal protein S7e signature

A number of eukaryotic ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian S7.

- *Xenopus* S8.

- Insect S7.

- Yeast probable ribosomal protein S7 (N2212).

- Fission yeast probable ribosomal protein S7 (SpAC18G6.13c).

These proteins have about 200 amino acids. A highly conserved stretch of 14 residues which is located in the central section and which is rich in charged residues was selected as a signature pattern.

Consensus pattern: [KR]-L-x-R-E-L-E-K-K-F-[SAP]-x-[KR]-H

[1] Salazar C.E., Mills-Hamm D.M., Kumar V., Collins F.H. *Nucleic Acids Res.* 21:4147-4147(1993).

5 565. Ribosomal protein S8 signature

Ribosomal protein S8 is one of the proteins from the small ribosomal subunit. In *Escherichia coli*, S8 is known to bind directly to 16S ribosomal RNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups: - Eubacterial S8, - Algal and plant chloroplast S8.

10 - Cyanelle S8. - Archaeobacterial S8. - *Marchantia polymorpha* mitochondrial S8. - Mammalian S15A. - Plant S15A. - Yeast S22 (S24).

The best conserved region located in the C-terminal section of these proteins has been selected as a signature pattern.

-Consensus pattern: [GE]-x(2)-[LIV](2)-[STY]-[ST]-x(2)-G-[LIVM](2)-x(4)-[AG]-
15 [KRHAYI]

[1] Otaka E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

20 566. Ribosomal protein S8e signature

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities [1]. One of these families consists of:

- Mammalian S8. - *Caenorhabditis elegans* S8 (F42C5.8). - *Leishmania major* S8. - Plant S8. - Yeast S8 (S14) (Rp19). - Archeobacterial S8e.

25 These proteins have either about 220 amino acids (in eukaryotes) or about 125 amino acids (in archaeobacteria). A conserved stretch which is located in the N-terminal section and which is rich in positively charged residues has been selected as a signature pattern.

-Consensus pattern: [KR]-x(2)-[ST]-G-[GA]-x(5)-[HR]-[KG]-[KR]-x-K-x-E-[LM]-G

30 [1] Engemann S., Herfurth E., Briesemeister U., Wittmann-Liebold B.

J. Protein Chem. 14:189-195(1995).

567. Ribosomal protein S9 signature

Ribosomal protein S9 is one of the proteins from the small ribosomal subunit.

It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1,2], groups: - Eubacterial S9. - Algal chloroplast S9.

- Cyanelle S9. - Archaeobacterial S9. - Mammalian S16. - Plant S16.

- Yeast mitochondrial ribosomal S9.

A conserved region containing many charged residues and located in the central section of these proteins has been selected as a signature pattern.

-Consensus pattern: G-G-G-x(2)-[GSA]-Q-x(2)-[SA]-x(3)-[GSA]-x-[GSTAV]-[KR]-

[GSA]-[LIF]

[1] Chan Y.-L., Paz V., Olvera J., Wool I.G. FEBS Lett. 263:85-88(1990).

[2] Otake E., Hashimoto T., Mizuta K.

Protein Seq. Data Anal. 5:285-300(1993).

568. Ribulose-phosphate 3-epimerase family signatures

Ribulose-phosphate 3-epimerase (EC 5.1.3.1) (also known as pentose-5-phosphate

3-epimerase or PPE) is the enzyme that converts D-ribulose 5-phosphate into

D-xylulose 5-phosphate in Calvin's reductive pentose phosphate cycle. In

Alcaligenes eutrophus two copies of the gene coding for PPE are known [1], one is chromosomally encoded (cbbEC), the other one is on a plasmid (cbbEP).

PPE has been found in a wide range of bacteria, archaebacteria, fungi and plants. The sequence of PPE is highly related to:

- *Escherichia coli* D-allulose-6-phosphate 3-epimerase (gene *alsE*).

- *Escherichia coli* protein *sgcE*.

- *Mycoplasma genitalium* hypothetical protein MG112.

All these proteins have from 209 to 241 amino acid residues.

Two conserved regions which are located respectively in the N-terminal and in the central part of these proteins have been selected as signature patterns.

-Consensus pattern: [LIVMF]-H-[LIVMFY]-D-[LIVM]-x-D-x(1,2)-[FY]-[LIVM]-x-N-x-[STAV]

-Consensus pattern: [LIVMA]-x-[LIVM]-M-[ST]-[VS]-x-P-x(3)-G-Q-x-F-x(6)-[NK]-[LIVMC]

[1] Kusian B., Yoo J.G., Bednarski R., Bowien B.
J. Bacteriol. 174:7337-7344(1992).

5 569. (Ricin B lectin) Similarity to lectin domain of ricin beta-chain, 3 copies.

This family consists of a triplicated domain involved in
cell agglutination in ricin.

10

570. (Rotamase) PpiC-type peptidyl-prolyl cis-trans isomerase signature
Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase or rotamase) is an
enzyme that accelerates protein folding by catalyzing the cis-trans
isomerization of proline imidic peptide bonds in oligopeptides [1]. Most
15 characterized PPIases belong to two families, the cyclophilin-type (see
<PDOC00154>) and the FKBP-type (see <PDOC00426>). Recently a third family
has been discovered [2,3]. So far, the only biochemically characterized
member of this family is the *Escherichia coli* protein parvulin (gene *ppiC*), a
small (92 residues) cytoplasmic enzyme that prefers amino acid residues with
20 hydrophobic side chains like leucine and phenylalanine in the P1 position of
the peptides substrates. PpiC is evolutionary related to a number of proteins
that are also probably PPIases:

- *Escherichia coli* and *Haemophilus influenzae* *ppiD*. PpiD is a PPIase which
contains a periplasmic *ppiC*-like domain anchored to the inner membrane and
25 which seems to be involved in the folding of outer membrane proteins.
- *Escherichia coli* *surA*. *SurA* is a periplasmic protein that contains two
ppiC-like domains.
- Nitrogen-assimilating bacteria protein *nifM* which is involved in the
activation and stabilization of the iron-component (*nifH*) of nitrogenase.
- 30 - *Bacillus subtilis* protein *prsA*, a membrane-bound lipoprotein involved in
protein export.
- *Lactococcus* and *Lactobacillus* protease maturation protein *prtM*, a membrane-
bound lipoprotein involved in the maturation of a secreted serine

proteinase. - Yeast protein ESS1/PTF1 (processing/termination factor 1).

- Drosophila protein dodo (gene dod). - Mammalian protein PIN1,

- Campylobacter jejuni cell binding factor 2 (CBF2), a secreted antigen.

- Bacillus subtilis hypothetical protein yacD.

5 - Helicobacter pylori hypothetical protein HP0175.

- A hypothetical slime mold protein.

A conserved region that contains a serine which could play a role in the catalytic mechanism of these enzymes has been selected as a signature pattern.

-Consensus pattern: F-[GSADEI]-x-[LVAQ]-A-x(3)-[ST]-x(3,4)-[STQ]-x(3,5)-[GER]-

10 G-x-[LIVM]-[GS]

[1] Fischer G., Schmid F.X.

Biochemistry 29:2205-2212(1990).

[2] Rudd K.E., Sofia H.J., Koonin E.V., Plunkett G. III, Lazar S.,

Rouviere P.E. Trends Biochem. Sci. 20:14-15(1995).

15 [3] Rahfeld J.-U., Ruecknagel K.P., Schelbert B., Ludwig B., Hacker J.,

Mann K., Fischer G. FEBS Lett. 352:180-184(1994).

571. (RnaAD) Ribosomal RNA adenine dimethylases signature

20 A number of enzymes responsible for the dimethylation of adenosines if ribosomal RNAs (EC 2.1.1.48) have been found [1,2] to be evolutionary related.

These enzymes are:

- Bacterial 16S rRNA dimethylase (gene ksgA), which acts in the biogenesis of ribosomes by catalyzing the dimethylation of two adjacent adenosines in

25 the loop of a conserved hairpin near the 3'-end of 16S rRNA. Inactivation of ksgA leads to resistance to the aminoglycoside antibiotic kasugamycin.

- Yeast 18S rRNA dimethylase (gene DIM1), which is functionally similar to ksgA and that dimethylates twin adenosines in the 3'-end of 18S rRNA.

- Bacterial 'erm' methylases. These enzymes confer resistance to macrolide-lincosamide-streptogramin B (MLS) antibiotics - such as erythromycin - by
30 dimethylating the adenine residue at position 2058 of 23S rRNA thus resulting in a reduced affinity between ribosomes and the MLS antibiotics.

- Caenorhabditis elegans hypothetical protein EO2H1.1.

The best conserved regions in these enzymes is located in the N-terminal section and corresponds to a region that is probably involved in S-adenosyl methionine (SAM) binding.

-Consensus pattern: [LIVM]-[LIVMFY]-[DE]-x-G-[STAPV]-G-x-[GA]-x-[LIVMF]-[ST]-
x(2)-[LIVM]-x(6)-[LIVMY]-x-[STAGV]-[LIVMFYHC]-E-x-D

[1] van Gemen B., van Knippenberg P.H.

(In) Nucleic acid methylation, Clawson G.A., Willis D.B., Weissbach A., Jones P.A., Eds., pp.19-36, Alan R. Liss Inc, New-York, (1990).

[2] Lafontaine D., Delcour J., Glasser A.L., Desgres J., Vandenhoute J.

J. Mol. Biol. 241:492-497(1994).

572. (RuBisCO small) Ribulose biphosphate carboxylase, small chain. 206 members

573. ATP/GTP-binding site motif A (P-loop) (ras)

From sequence comparisons and crystallographic data analysis it has been shown

[1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix.

This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5]. There are numerous

ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such a motif has been noted are listed below: - ATP

synthase alpha and beta subunits. - Myosin heavy chains. - Kinesin heavy chains and kinesin-like proteins. - Dynamins and dynamin-like proteins - Guanylate kinase - Thymidine kinase (- Thymidylate kinase. - Shikimate kinase. - Nitrogenase iron protein family (nifH/frxC) - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] - DNA and RNA

helicases [8,9,10]. - GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.). -

Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.). - Nuclear protein ran. - ADP-ribosylation factors family - Bacterial dnaA protein - Bacterial recA protein - Bacterial recF protein - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.). - DNA mismatch repair proteins mutS family - Bacterial type II secretion system

In addition to the proteins listed above, the 'A' motif is also found in a number of other proteins. Most of these proteins probably bind a nucleotide, but others are definitively not ATP- or GTP-binding (as for example chymotrypsin, or human ferritin light chain).

GTP-binding nuclear protein ran signature (ras)

Ran (or TC4) is a small abundant nuclear protein that binds and hydrolyzes GTP and which has been implicated in a large number of processes including nucleocytoplasmic transport, RNA synthesis, processing and export and cell cycle checkpoint control [1,2]. Ran is generally included in the RAS 'superfamily' of small GTP-binding proteins [3], but it is only slightly related to the other RAS proteins. It also differs from RAS proteins in that it lacks cysteine residues at its C- terminal and is therefore not subject to prenylation. Instead ran has an acidic C-terminus. It is, however similar to RAS family members in requiring a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. The region of the GTP-binding B motif which, in ran, is perfectly conserved has been selected as a signature pattern.

Consensus pattern: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y- Proteins belonging to this family also contain a copy of the ATP/GTP- binding motif 'A' (P-loop).

[1] Scheffzek K., Klebe C., Fritz-Wolf K., Kabsch W., Wittinghofer A. Nature 374:378-381(1995).[2] Rush M.G., Drivas G., d'Eustachio P. BioEssays 18:103-112(1996).[3]

5 Valencia A., Chardin P., Wittinghofer A., Sander C. Biochemistry 30:4637-4648(1991).

574. recA signature

The bacterial recA protein [1,2,3,E1] is essential for homologous recombination and recombinational repair of DNA damage. RecA has many activities: it filaments, it binds to single- and double-stranded DNA, it binds and hydrolyzes ATP, it is also a recombinase and, finally, it interacts with lexA causing its activation and leading to its autocatalytic cleavage. RecA is a protein of about 350 amino-acid residues. Its sequence is very well conserved [3,4,5,E1] among eubacterial species. It is also found in the chloroplast of plants [6]. The best conserved region, a nonapeptide located in the middle of the sequence which is part of the monomer-monomer interface in a recA filament has been selected as a signature pattern., Consensus pattern: A-L-[KR]-[IF]-[FY]-[STA]-[STAD]-[LIVMQ]-R-

[1] Smith K.C., Wang T.-C. V. BioEssays 10:12-16(1989).[2] Lloyd A.T., Sharp P.M. J. Mol. Evol. 37:399-407(1993).[3] Roca A.L., Cox M.M. Prog. Nucleic Acids Res. Mol. Biol. 56:129-223(1997).[4] Karlin S., Weinstock G.M., Brendel V. J. Bacteriol. 177:6881-6893(1995).[5] Eisen J.A. J. Mol. Evol. 41:1105-1123(1995).[6] Cerutti H.D., Osman M., Grandoni P., Jagendorf A.T. Proc. Natl. Acad. Sci. U.S.A. 89:8068-8072(1992).[E1] <http://www.tigr.org/~jeisen/RecA/RecA.html>

575. Response regulator receiver domain

This domain receives the signal from the sensor partner inComment: bacterial two-component systems. It is usually found N-terminalComment: to a DNA binding effector domain.

[1] Pao GM, Saier MH; J Mol Evol 1995;40:136-154.

576. Ribonucleotide reductase large subunit signature

*Ribonucleotide reductase (EC 1.17.4.1) [1,2] catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. Ribonucleotide reductase is an oligomeric enzyme composed of a large subunit (700 to 1000 residues) and a small subunit (300 to 400 residues). There are regions of similarities in the sequence of the large chain from prokaryotes, eukaryotes and viruses. One of these regions has been selected as a signature pattern.

Consensus pattern: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-[PA]-

[1] Nillson O., Lundqvist T., Hahne S., Sjöberg B.-M. Biochem. Soc. Trans. 16:91-94(1988). [2] Reichard P. Science 260:1773-1777(1993).

577. Ribonuclease T2 family histidine active sites

The fungal ribonucleases T2 from *Aspergillus oryzae*, M from *Aspergillus saitoi* and Rh from *Rhizopus niveus* are structurally and functionally related 30 Kd glycoproteins [1] that cleave the 3'-5' internucleotide linkage of RNA via a nucleotide 2',3'-cyclic phosphate intermediates (EC 3.1.27.1). A number of other RNases have been found to be evolutionarily related to these fungal enzymes: - Self-incompatibility [2] in flowering plants is often controlled by a single gene (S-gene) that has several alleles. This gene prevents fertilization by self-pollen or by pollen bearing either of the two S-alleles expressed in the style. The self-incompatibility glycoprotein from several higher plants of the solanaceae family has been shown [2,3] to be a ribonuclease. - Phosphate-starvation induced RNases LE and LX from tomato [4]. These two enzymes are probably involved in a phosphate-starvation rescue system. - *Escherichia coli* periplasmic RNase I (EC 3.1.27.6) (gene *rna*) [5]. - *Aeromonas hydrophila* periplasmic RNase. - *Haemophilus influenzae* hypothetical protein HI0526. Two histidine residues have been shown [6,7] to be involved in the catalytic mechanism of RNase T2 and Rh. These residues and the region around them are highly conserved in all the sequence described above. Two signature patterns have been developed, one for each of the two active-site histidines. The second pattern also contains a cysteine which is known to be involved in a disulfide bond.

Consensus pattern: [FYWL]-x-[LIVM]-H-G-L-W-P [H is an active site residue]

Consensus pattern: [LIVMF]-x(2)-[HDGTY]-[EQ]-[FYW]-x-[KR]-H-G-x-C [H is an active site residue] [C is involved in a disulfide bond]

[1] Watanabe H., Naitoh A., Suyama Y., Inokuchi N., Shimada H., Koyama T., Ohgi K., Irie M. J. Biochem. 108:303-310(1990).[2] Haring V., Gray J.E., McClure B.A., Anderson M.A., Clarke A.E. Science 250:937-941(1990).[3] McClure B.A., Haring V., Ebert P.R., Anderson M.A., Simpson R.J., Sakiyama F., Clarke A.E. Nature 342:95957(1989).[4] Loeffler A., Glund K., Irie M. Eur. J. Biochem. 214:627-633(1993).[5] Meador J. III, Kennell D. Gene 95:1-7(1990).[6] Kawata Y., Sakiyama F., Hayashi F., Kyogoku Y. Eur. J. Biochem. 187:255-262(1990).[7] Kurihara H., Mitsui Y., Ohgi K., Irie M., Mizuno H., Nakamura K.T. FEBS Lett. 306:189-192(1992).

578. Ribonucleotide reductase large subunit signature. Ribonucleotide reductase (EC 1.17.4.1) [1,2] catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. Ribonucleotide reductase is an oligomeric enzyme composed of a large subunit (700 to 1000 residues) and a small subunit (300 to 400 residues). There are regions of similarities in the sequence of the large chain from prokaryotes, eukaryotes and viruses. One of these regions has been developed as a signature pattern.

Consensus pattern: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-[PA]-

[1] Nillson O., Lundqvist T., Hahne S., Sjöberg B.-M. Biochem. Soc. Trans. 16:91-94(1988).[2] Reichard P. Science 260:1773-1777(1993).

579. RNase H

RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers.

580. Eukaryotic putative RNA-binding region RNP-1 signature (rrm)

Many eukaryotic proteins that are known or supposed to bind single-stranded RNA contain one or more copies of a putative RNA-binding domain of about 90 amino acids [1,2]. This region has been found in the following proteins: ** Heterogeneous nuclear ribonucleoproteins ** - hnRNP A1 (helix destabilizing protein) (twice). - hnRNP A2/B1 (twice). - hnRNP C (C1/C2) (once). - hnRNP E (UP2) (at least once). - hnRNP G (once). ** Small nuclear ribonucleoproteins ** - U1 snRNP 70 Kd (once). - U1 snRNP A (once). - U2 snRNP B" (once). ** Pre-RNA and mRNA associated proteins ** - Protein synthesis initiation factor 4B (eIF-4B) [3], a protein essential for the binding of mRNA to ribosomes (once). - Nucleolin (4 times). - Yeast single-stranded nucleic acid-binding protein (gene SSB1) (once). - Yeast protein NSR1 (twice). NSR1 is involved in pre-rRNA processing; it specifically binds nuclear localization sequences. - Poly(A) binding protein (PABP) (4 times). ** Others ** - Drosophila sex determination protein Sex-lethal (Sxl) (twice). - Drosophila sex determination protein Transformer-2 (Tra-2) (once). - Drosophila 'elav' protein (3 times), which is probably involved in the RNA metabolism of neurons. - Human paraneoplastic encephalomyelitis antigen HuD (3 times) [4], which is highly similar to elav and which may play a role in neuron-specific RNA processing. - Drosophila 'bicoid' protein (once) [5], a segment-polarity homeobox protein that may also bind to specific mRNAs. - La antigen (once), a protein which may play a role in the transcription of RNA polymerase III. - The 60 Kd Ro protein (once), a putative RNP complex protein. - A maize protein induced by abscisic acid in response to water stress, which seems to be a RNA-binding protein. - Three tobacco proteins, located in the chloroplast [6], which may be involved in splicing and/or processing of chloroplast RNAs (twice). - X16 [7], a mammalian protein which may be involved in RNA processing in relation with cellular proliferation and/or maturation. - Insulin-induced growth response protein Cl-4 from rat (twice). - Nucleolysins TIA-1 and TIAR (3 times) [8] which possesses nucleolytic activity against cytotoxic lymphocyte target cells, may be involved in apoptosis. - Yeast RNA15 protein, which plays a role in mRNA stability and/or poly-(A) tail length [9]. Inside the putative RNA-binding domain there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif), the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation:

xxxxxxxx#####xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx#####xxxxxxxxxxxxxxxxxxxxxxxxxxxx
RNP-2 RNP-1

The RNP-1 motif has been used as a signature pattern for this type of domain.

- 5 Consensus pattern: [RK]-G-{EDRKHPG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM] In most cases the residue in position 3 of the pattern is either Tyr or Phe.

- [1] Bandziulis R.J., Swanson M.S., Dreyfuss G. *Genes Dev.* 3:431-437(1989).[2] Dreyfuss G., Swanson M.S., Pinol-Roma S. *Trends Biochem. Sci.* 13:86-91(1988).[3] Milburn S.C., Hershey J.W.B., Davies M.V., Kelleher K., Kaufman R.J. *EMBO J.* 9:2783-2790(1990).[4] Szabo A., Dalmau J., Manley G., Rosenfeld M., Wong E., Henson J., Posner J.B., Furneaux H.M. *Cell* 67:325-333(1991).[5] Rebagliati M. *Cell* 58:231-232(1989).[6] Li Y., Sugiura M. *EMBO J.* 9:3059-3066(1990).[7] Ayane M., Preuss U., Koehler G., Nielsen P.J. *Nucleic Acids Res.* 19:1273-1278(1991).[8] Kawakami A., Tian Q., Duan X., Streuli M., Schlossman S.F., Anderson P. *Proc. Natl. Acad. Sci. U.S.A.* 89:8681-8685(1992).[9] Minvielle-Sebastia L., Winsor B., Bonneaud N., Lacroute F. *Mol. Cell. Biol.* 11:3075-3087(1991).

581. Rubredoxin signature

Rubredoxins [1] are small electron-transfer prokaryotic proteins. They contain an iron atom which is ligated by four cysteine residues. Rubredoxins are, in some cases, functionally interchangeable with ferredoxins.

A conserved region that includes two of the cysteine residues that bind the iron atom has been selected as a pattern for these proteins.

25 Consensus pattern: [LIVM]-x(3)-W-x-C-P-x-C-[AGD] [The two C's bind the iron atom]

In *Pseudomonas oleovorans* rubredoxin 2 (gene *alkG*) [2], this pattern is found twice because *alkG* has two rubredoxin domains.

Rubrerithrin [3], a protein with inorganic pyrophosphatase activity from *Desulfovibrio vulgaris* possesses a C-terminal rubredoxin-like domain, but this domain is too divergent to be detected by the above pattern.

[1] Berg J.M., Holm R.H.(In) *Iron-sulfur proteins*, Spiro T.G., Ed., pp1-66, Wiley, New-York, (1982). [2] Kok M., Oldenhuis R., der Linden M.P.G., Meulenberg C.H.C.,

Kingma J., Witholt B., *J. Biol. Chem.* 264:5442-5451(1989). [3] van Beeumen J.J., van Driessche G., Liu M.-Y., Le Gall J., *J. Biol. Chem.* 266:20645-20653(1991).

582. (rvp) Eukaryotic and viral aspartyl proteases active site
- Aspartyl proteases, also known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes [1,2,3] known to exist invertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases are: -
- Vertebrate gastric pepsins A and C (also known as gastricsin). - Vertebrate chymosin (rennin), involved in digestion and used for making cheese. - Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34). - Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma. - Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21). - Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases. -
- Yeast barrier pepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone. - Fission yeast *ssa1* which is involved in degrading or processing the mating pheromones. Most retroviruses and some plant viruses, such as badnaviruses, encode for an aspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of a polypeptide which is cleaved during the maturation process of the virus. It is generally part of the polypeptide and, more rarely, of the gagpolypeptide. Conservation of the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases allows us to develop a single signature pattern for both groups of protease.
- Consensus pattern: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-x-[LIVMFSTNC]-x-[LIVMFGTA] [D is the active site residue] -
- [1] Foltmann B. *Essays Biochem.* 17:52-84(1981). [2] Davies D.R. *Annu. Rev. Biophys. Chem.* 19:189-215(1990). [3] Rao J.K.M., Erickson J.W., Wlodawer A. *Biochemistry* 30:4663-4671(1991). [4] Rawlings N.D., Barrett A.J. *Meth. Enzymol.* 248:105-120(1995).

583. (rvt) Reverse transcriptase (RNA-dependent DNA polymerase)

A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses. Number of members: 1233

[1] Medline: 91006031. Origin and evolution of retroelements based upon their reverse transcriptase sequences. Xiong Y, Eickbush TH; EMBO J 1990;9:3353-3362.

584. (S-AdoMet synt) S-adenosylmethionine synthetase signatures

S-adenosylmethionine synthetase (EC 2.5.1.6) is the enzyme that catalyzes the formation of S-adenosylmethionine (AdoMet) from methionine and ATP [1]. AdoMet is an important methyl donor for transmethylation and is also the propylamino donor in polyamine biosynthesis. In bacteria there is a single isoform of AdoMet synthetase (gene metK), there are two in budding yeast (genes SAM1 and SAM2) and in mammals while in plants there is generally a multigene family. The sequence of AdoMet synthetase is highly conserved throughout isoforms and species. Two signature patterns have been selected for this type of enzyme; the first is a hexapeptide which seems to be involved in ATP-binding; the second is an almost perfectly conserved glycine-rich nonapeptide.

Consensus pattern: G-A-G-D-Q-G-x(3)-G-[FYH]-Sequences known to belong to this class detected by the pattern:

Consensus pattern: G-[GA]-G-[ASC]-F-S-x-K-[DE]

[1] Horikawa S., Sasuga J., Shimizu K., Ozasa H., Tsukada K. J. Biol. Chem. 265:13683-13686(1990).

585. S1 RNA binding domain

The S1 domain occurs in a wide range of RNA-associated proteins. It is structurally similar to cold shock protein which binds nucleic acids. Comment: The S1 domain has an OB-fold structure.

[1] Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG; Cell 1997;88:235-242.

586. SAICAR synthetase signatures

Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6) (SAICARSynthetase) catalyzes the seventh step in the de novo purine biosynthetic pathway; the ATP-dependent conversion of 5'-phosphoribosyl-5-aminoimidazole-4-carboxylic acid and aspartic acid to SAICAR [1]. In bacteria (gene *purC*), fungi (gene *ADE1*) and plants, SAICAR synthetase is a monofunctional protein; in higher vertebrates it is the N-terminal domain of a bifunctional enzyme that also catalyze phosphoribosylaminoimidazole carboxylase (AIRC) activity. Two conserved regions in the central section of this enzyme have been selected as signature patterns for SAICAR synthetase.

Consensus pattern: [LIVMF](2)-P-[LIVM]-E-x-[LIVM]-[LIVMCA]-R-x(3)-[TA]-G-S-
 Consensus pattern: [LIVM]-[LIVMA]-D-x-K-[LIVMFY]-E-F-G
 [1] Zalkin H., Dixon J.E. Prog. Nucleic Acid Res. Mol. Biol. 42:259-287(1992).

587. (SCP) Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signatures

A variety of extracellular proteins from eukaryotes have been found to be evolutionary related: - Rodent sperm-coating glycoprotein (SCP), also known as acidic epididymal glycoprotein (AEG). This protein is thought to be involved in sperm maturation [1]. It is a protein of about 220 residues and probably contains eight disulfide bonds. - Mammalian testis-specific protein Tpx-1 [2]. Tpx-1 is highly related to SCP's. - Mammalian glioma pathogenesis-related protein (GliPR). - Lizard helothermine, a toxin that blocks ryanodine receptors. - Venom allergen 5 (Ag5) from vespid wasps and venom allergen 3 (Ag3) from fire ants. These proteins are potent allergens and are the main cause of allergic reactions to stings from insects of the hymenoptera family [3]. Ag5/3 are proteins of about 200 residues and contain four disulfide bonds. - Plant pathogenesis proteins of the PR-1 family [4]. These proteins are synthesized during pathogen infection or other stress-related responses. They are proteins of about 130 to 140 residues and probably contain three disulfide bonds. - Proteins Sc7 and Sc14 from the basidiomycete fungus *Schizophyllum commune*. These extracellular proteins are loosely associated with fruit body hyphal walls [5]. Sc7/14 are proteins of about 180 residues and probably contain two disulfide bonds. - *Ancylostoma* secreted protein from

dog hookworm. - Yeast hypothetical proteins YJL078c, YJL079c and YKR013w. The exact function of these proteins is not yet known. Two conserved regions located in their C-terminal half have been selected as signature patterns. The second signature contains a cysteine which is known to be involved in a disulfide bond in Ag5.

5 Consensus pattern: [GDER]-H-[FYWH]-T-Q-[LIVM](2)-W-x(2)-[STN]

Consensus pattern: [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N-[LIVMFYWDN] [C is involved in a disulfide bond]

[1] Mizuki N., Kasahara M. Mol. Cell. Endocrinol. 89:25-32(1992).[2] Kasahara M., Gutknecht J., Brew K., Spurr N., Goodfellow P.N. Genomics 5:527-534(1989).[3] Lu G.,

10 Villalba M., Coscia M.R., Hoffman D.R., King T.P. J. Immunol. 150:2823-2830(1993).[4]

Dixon D.C., Cutt J.R., Klessig D.F. EMBO J. 10:1317-1324(1991).[5] Schuren F.H.J., Asgeirsdottir S.A., Kothe E.M., Scheer J.M.J., Wessels J.G.H. J. Gen. Microbiol. 139:2083-2090(1993).

15 588. SET domain

SET domains appear to be protein-protein interactionComment: domains. It has been demonstrated that SET domainsComment: mediate interactions with a family of proteins thatComment: display similarity with dual-specificity phosphatasesComment: (dsPTases) [2].

[1] Tripoulas N, LaJeunesse D, Gildea J, Shearn A; Genetics 1996;143:913-928. [2] Cui X, De Vivo I, Slany R, Miyamoto A, Firestein R, Cleary, ML; Nat Genet 1998;18:331-337.

25 589. Src homology 3 (SH3) domain profile

The Src homology 3 (SH3) domain is a small protein domain of about 60 amino-acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g. Src, Abl, Lck) [1]. Since then, it has been found in a great variety of other intracellular or membrane-associated proteins [2,3,4,5]. The SH3 domain has a characteristic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices [6]. The function of the SH3 domain is not well understood. The current opinion is that they mediate assembly of specific protein complexes via binding to proline-rich peptides [7]. In general SH3 domains

are found as single copies in a given protein, but there is a significant number of protein with two SH3 domains and a few with 3 or 4 copies. So far, SH3 domains have been identified in the following proteins: - Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases. - Mammalian phosphatidylinositol-specific phospholipase C-gamma-1 and -2. - Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit. - Mammalian Ras GTPase-activating protein (GAP). - Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, *Caenorhabditis elegans* sem-5 and *Drosophila* DRK. All of which have two SH3 domains. - Mammalian Vav oncoprotein, a guanine nucleotide exchange factor of the CDC24 family. - Some guanine-nucleotide releasing factors of the CDC25 family: yeast CDC25, yeast SCD25, fission yeast ste6. - MAGUK proteins. These proteins consist of at least three types of domains: one or more copies of the DHR domain, a SH3 domain and a C-terminal guanylate kinase domain. Members of this family are: *Drosophila* lethal(1) discs large-1 tumor suppressor protein (gene Dlg1), mammalian tight junction protein ZO-1, vertebrate erythrocyte membrane protein p55, *Caenorhabditis elegans* protein lin-2, rat protein CASK and mammalian synaptic proteins SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1 and SAP102. - Miscellaneous proteins interacting with vertebrate receptor protein tyrosine kinases: mammalian cytoplasmic protein Nck (3 copies), oncoprotein Crk (2 copies). - Chicken Src substrate p80/85 protein (cortactin) and the similar human hemopoietic lineage cell specific protein Hs1. - Mammalian dihydropyridine-sensitive L-type calcium channel beta (regulatory) subunit including the related human myasthenic syndrome antigen B (MSYB). - Mammalian neutrophil cytosolic activators of NADPH oxidase: p47 (NCF-1), p67 (NCF-2), and a potential homolog from *Caenorhabditis elegans* (B0303.7). NCF-1 and -2 have two copies of the SH3 domain, while B0303.7 has four. - Some myosin heavy chains from amoebae, slime molds and yeast (gene MYO3). - Vertebrate and *Drosophila* spectrin and fodrin alpha-chain. - Human amphiphysin. - Yeast actin-binding protein ABP1. - Yeast actin-binding protein SLA1 (3 copies). - Yeast protein BEM1 and the fission yeast homolog scd2 (or ra3) (2 copies). - Yeast BEM1-binding proteins BOI2 (BEB1) and BOB1 (BOI1). - Yeast fusion protein FUS1. - Yeast protein RSV167. - Yeast protein SSU81. - Yeast hypothetical proteins YAR014c (1 copy), YFR024c (1 copy), YHL002w (1 copy), YHR016c (1 copy), YJL020C (1 copy), YHR114w (2 copies) and the fission yeast homolog SpAC12C2.05c. - *Caenorhabditis elegans* hypothetical proteins F42H10.3. The profile developed to detect SH3

domains is based on a structural alignment consisting of 5 gap-free blocks and 4 linker regions totaling 62 match positions.

[1] Mayer B.J., Hamaguchi M., Hanafusa H. *Nature* 332:272-275(1988).[2] Musacchio A., Gibson T., Lehto V.P., Saraste M. *FEBS Lett.* 307:55-61(1992).[3] Pawson T., Schlessinger J. *Curr. Biol.* 3:434-442(1993).[4] Mayer B.J., Baltimore D. *Trends Cell Biol.* 3:8-13(1993).[5] Pawson T. *Nature* 373:573-580(1995).[6] Kuriyan J., Cowburn D. *Curr. Opin. Struct. Biol.* 3:828-837(1993).[7] Morton C.J., Campbell I.D. *Curr. Biol.* 4:615-617(1994).

590. Serine hydroxymethyltransferase pyridoxal-phosphate attachment site (SHMT)
Serine hydroxymethyltransferase (EC 2.1.2.1) (SHMT) [1] catalyzes the transfer of the hydroxymethyl group of serine to tetrahydrofolate to form 5,10-methylenetetrahydrofolate and glycine. In vertebrates, it exists in acytoplasmic and a mitochondrial form whereas only one form is found in prokaryotes. Serine hydroxymethyltransferase is a pyridoxal-phosphate containing enzyme. The pyridoxal-P group is attached to a lysine residue around which the sequence is highly conserved in all forms of the enzyme.

Consensus pattern: [DEH]-[LIVMFY]-x-[STMV]-[GST]-[ST](2)-H-K-[ST]-[LF]-x-G-[PAC]-[RQ]-[GSA]-[GA] [K is the pyridoxal-P attachment site]

[1] Usha R., Savithri H.S., Rao N.A. *Biochim. Biophys. Acta* 1204:75-83(1994).

591. SIS domain

SIS (Sugar ISomerase) domains are found in many phosphosugar isomerases and phosphosugar binding proteins.

[1] Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B, Polikarpov I; *Structure* 1998;6:1047-1055.

592. (SKI) Shikimate kinase signature

Shikimate kinase (EC 2.7.1.71) catalyzes the fifth step in the biosynthesis from chorismate of the aromatic amino acids (the shikimate pathway) in bacteria (gene *aroK* or *aroL*), plants and in fungi (where it is part of a multifunctional enzyme which catalyzes five consecutive steps

493

in this pathway). Shikimate kinase is a small protein of about 200 residues. A conserved region that contains a run of three glycines has been selected as a signature pattern.

Consensus pattern: [KR]-x(2)-E-x(3)-[LIVMF]-x(8,12)-[LIVMF](2)-[SA]-x-G(3)-x-[LIVMF]. Proteins belonging to this family also contain a copy of the ATP/GTP-binding motif 'A' (P-loop).

593. SNAP-25 family

SNAP-25 (synaptosome-associated protein 25 kDa) proteins are components of SNARE complexes. Members of this family contain a cluster of cysteine residues that can be palmitoylated for membrane attachment [2].

[1] Brennwald P, Kearns B, Champion K, Keranen S, Bankaitis V, Novick P; Cell 1994;79:245-258. [2] Risinger C, Blomqvist AG, Lundell I, Lambertsson A, Nassel D, Pieribone VA, Brodin L, Larhammar D; J Biol Chem 1993;268:24408-24414.

594. SNF2 and others N-terminal domain

This domain is found in proteins involved in a variety of processes including transcription regulation (e.g., SNF2, STH1, brahma, MOT1), DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g., Iodestar, ETL1).

595. Staphylococcal nuclease homologues (Snase)

Present in all three domains of cellular life. Four copies in the transcriptional coactivator p100. These, however, appear to lack the active site residues of Staphylococcal nuclease.

Positions 14 (Asp-21), 34 (Arg-35), 39 (Asp-40), 42 (Glu-43) and Comment: 110 (Arg-87) [Snase numbering in parentheses] are thought to be involved in substrate-binding and catalysis.

[1] Ponting CP; Protein Sci 1997;6:459-463. [2] Callebaut I, Mornon JP; Biochem J 1997;321:125-132.

596. SPRY domainA

SPRY Domain is named from SPLa and the RYanodine Receptor. Domain of unknown function. Distant homologues are domains in Comment: butyrophilin/marenostrin/pyrin homologues.

[1] Ponting C, Schultz J, Bork P; Trends Biochem Sci 1997;22:193-194.

597. (SQS PSY) Squalene and phytoene synthases signatures

Two different polyisoprene synthases have been shown [1,2,3] to share a number of regions of sequence similarities: - Squalene synthase (EC 2.5.1.21) (farnesyl-diphosphate farnesyltransferase) (SQS), which catalyzes the conversion of two molecules of farnesyl diphosphate (FPP) into squalene. It is the first committed step in the cholesterol biosynthetic pathway. The reaction carried out by SQS is catalyzed in two separate steps: the first is a head-to-head condensation of the two molecules of FPP to form presqualene diphosphate; this intermediate is then rearranged in a NADP-dependent reduction, to form squalene. SQS is found in eukaryotes. In yeast it is encoded by the ERG9 gene, in mammals by the FDFT1 gene. SQS seems to be membrane-bound. - Phytoene synthase (EC 2.5.1.-) (PSY), which catalyzes the conversion of two molecules of geranylgeranyl diphosphate (GGPP) into phytoene. It is the second step in the biosynthesis of carotenoids from isopentenyl diphosphate. The reaction carried out by PSY is catalyzed in two separate steps: the first is a head-to-head condensation of the two molecules of GGPP to form prephytoene diphosphate; this intermediate is then rearranged to form phytoene. PSY is found in all organisms that synthesize carotenoids: plants and photosynthetic bacteria as well as some non-photosynthetic bacteria and fungi. In bacteria PSY is encoded by the gene crtB. In plants PSY is localized in the chloroplast. As it can be seen from the description above, both SQS and PSY share a number of functional similarities which are also reflected at the level of their primary structure. In particular three well conserved regions are shared by SQS and PSY; they could be involved in substrate binding and/or the catalytic mechanism. Signature patterns

have been developed for the second and third conserved regions; they are localized in the central part of these enzymes.

Consensus pattern: Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]- x(2)-[LIV]

5 Consensus pattern: [LIVM]-G-x(3)-Q-x(2,3)-N-[IF]-x-R-D-[LIVMFY]-x(2)-[DE]- x(4,7)-R-x-[FY]-x-P-

[1] Summers C., Karst F., Charles A.D. Gene 136:185-192(1993).[2] Robinson G.W., Tsay Y.H., Kienzle B.K., Smith-Monroy C.A., Bishop R.W. Mol. Cell. Biol. 13:2706-

2727(1993).[3] Roemer S., Huguency P., Bouvier F., Camara B., Kuntz M. Biochem.

10 Biophys. Res. Commun. 196:1414-1421(1993).

598. SRP54-type proteins GTP-binding domain signature

The signal recognition particle (SRP) is an oligomeric complex that mediates targeting and insertion of the signal sequence of exported proteins into the membrane of the endoplasmic reticulum. SRP consists of a 7S RNA and six protein subunits. One of these subunits, the 54

Kd protein (SRP54), is a GTP-binding protein that interacts with the signal sequence when it emerges from the ribosome. The N-terminal 300 residues of SRP54 include the GTP-binding site (G-domain) and are evolutionary related to similar domains in other proteins which are

listed below [1]. - *Escherichia coli* and *Bacillus subtilis* ffh protein (P48), a protein which seems to be the prokaryotic counterpart of SRP54. Ffh is associated with a 4.5S RNA in the prokaryotic SRP complex. - Signal recognition particle receptor alpha subunit (docking

protein), an integral membrane GTP-binding protein which ensures, in conjunction with SRP, the correct targeting of nascent secretory proteins to the endoplasmic reticulum membrane.

The G-domain is located at the C-terminal extremity of the protein. - Bacterial ftsY protein, a protein which is believed to play a similar role to that of the docking protein in eukaryotes. The G-domain is located at the C-terminal extremity of the protein. - The pilA protein from *Neisseria gonorrhoeae* which seems to be the homolog of ftsY. - A protein from the

archaeobacteria *Sulfolobus solfataricus*. This protein is also believed to be a docking protein. The G-domain is also at the C-terminus. - Bacterial flagellar biosynthesis protein flhF. The best conserved regions in those domains are the sequence motifs that are part of the GTP-binding site, but as those regions are not specific to these proteins, they were not used as a

signature pattern. Instead, a conserved region located at the C-terminal end of the domain was selected.

Consensus pattern: P-[LIVM]-x-[FYL]-[LIVMAT]-[GS]-x-[GS]-[EQ]-x(4)-[LIVMF]
[1] Althoff S., Selinger D., Wise J.A. Nucleic Acids Res. 22:1933-1947(1994).

5

599. (STphosphatase) Serine/threonine specific protein phosphatases signature

Serine/threonine specific protein phosphatases (EC 3.1.3.16) (PP) [1,2,3] are enzymes that catalyze the removal of a phosphate group attached to a serine or evolutionary related. -

10 Protein phosphatase-1 (PP1) is an enzyme of broad specificity. It is inhibited by two thermostable proteins, inhibitor-1 and -2. In mammals, there are two closely related isoforms of PP-1: PP-1alpha and PP-1beta, produced by alternative splicing of the same gene. In

15 *Emericella nidulans*, PP-1 (gene bimG) plays an important role in mitosis control by reversing the action of the nimA kinase. In yeast, PP-1 (gene SIT4) is involved in dephosphorylating the large subunit of RNA polymerase II. - Protein phosphatase-2A (PP2A)

is also an enzyme of broad specificity. PP2A is a trimeric enzyme that consist of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit and a third variable subunit. In mammals, there are two closely related isoforms of the catalytic subunit of PP2A: PP2A-alpha and PP2A-beta, encoded by separate genes. - Protein phosphatase-2B (PP2B or calcineurin), a calcium-dependent enzyme whose activity is stimulated by

20 calmodulin. It is composed of two subunits: the catalytic A-subunit and the calcium-binding B-subunit. The specificity of PP2B is restricted. In addition to the above-mentioned enzymes, some additional serine/threonine specific protein phosphatases have been characterized and are listed below. - Mammalian phosphatase-X (PP-X), and *Drosophila* phosphatase-V (PP-V)

25 which are closely related but yet distinct from PP2A. - Yeast phosphatase PPH3, which is similar to PP2A, but with different enzymatic properties. - *Drosophila* phosphatase-Y (PP-Y), and yeast phosphatases Z1 and Z2 (genes PPZ1 and PPZ2) which are closely related but yet distinct from PP1. - *Drosophila* retinal degeneration protein C (gene rdgC), a calcium-binding phosphatase required to prevent light-induced retinal degeneration. - Phages Lambda and Phi-

30 80 ORF-221 which have been shown to have phosphatase activity and are related to mammalian PP's. The best conserved regions in these proteins is a perfectly conserved pentapeptide that can be used as a signature pattern.

Consensus pattern: [LIVM]-R-G-N-H-E-

[1] Cohen P. Annu. Rev. Biochem. 58:453-508(1989).[2] Cohen P., Cohen P.T.W. J. Biol. Chem. 264:21435-21438(1989).[3] Cohen P.T.W., Brewis N.D., Hughes V., Mann D.J. FEBS Lett. 268:355-359(1990).

5

600. Translation initiation factor SUI1 signature

In budding yeast (*Saccharomyces cerevisiae*), SUI1 is a translation initiation factor that functions in concert with eIF-2 and the initiator tRNA-Met in directing the ribosome to the proper start site of translation [1]. SUI1 is a protein of 108 residues. Close homologs of SUI1 have been found [2] in mammals, insects and plants. SUI1 is also evolutionary related to hypothetical proteins from *Escherichia coli* (yciH), *Haemophilus influenzae* (HI1225) and *Methanococcus vannielii*. A conserved region in the C-terminal section has been selected as a signature pattern.

Consensus pattern: [LIVM]-[EQ]-[LIVM]-Q-G-[DEN]-[KHQ]-[KRV]

[1] Yoon H., Donahue T.F. Mol. Cell. Biol. 12:248-260(1992).[2] Fields C.A., Adams M.D. Biochem. Biophys. Res. Commun. 198:288-291(1994).

601. (S T dehydratase) Serine/threonine dehydratases pyridoxal-phosphate attachment site

Serine and threonine dehydratases [1,2] are functionally and structurally related pyridoxal-phosphate dependent enzymes: - L-serine dehydratase (EC 4.2.1.13) and D-serine dehydratase (EC 4.2.1.14) catalyze the dehydration of L-serine (respectively D-serine) into ammonia and pyruvate. - Threonine dehydratase (EC 4.2.1.16) (TDH) catalyzes the dehydration of threonine into alpha-ketobuturate and ammonia. In *Escherichia coli* and other microorganisms, two classes of TDH are known to exist. One is involved in the biosynthesis of isoleucine, the other in hydroxamino acid catabolism. Threonine synthase (EC 4.2.99.2) is also a pyridoxal-phosphate enzyme, it catalyzes the transformation of homoserine-phosphate into threonine. It has been shown [3] that threonine synthase is distantly related to the serine/threonine dehydratases. In all these enzymes, the pyridoxal-phosphate group is attached to a lysine residue. The sequence around this residue is sufficiently conserved to allow the derivation of a pattern specific to serine/threonine dehydratases and threonine synthases.

Consensus pattern: [DESH]-x(4,5)-[STVG]-x-[AS]-[FYI]-K-[DLIFSA]-[RVMF]-[GA]-[LIVMGA] [The K is the pyridoxal-P attachment site]

[1] Ogawa H., Gomi T., Konishi K., Date T., Naakashima H., Nose K., Matsuda Y., Peraino C., Pitot H.C., Fujioka M. J. Biol. Chem. 264:15818-15823(1989).[2] Datta P., Goss T.J., Omnaas J.R., Patil R.V. Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).[3] Parsot C. EMBO J. 5:3013-3019(1986).[4] Grabowski R., Hofmeister A.E.M., Buckel W. Trends Biochem. Sci. 18:297-300(1993).

Cysteine synthase/cystathionine beta-synthase P-phosphate attachment site

Cysteine synthase (CSase) is the pyridoxal-phosphate dependent enzyme responsible [1] for the formation of cysteine from O-acetyl-serine and hydrogen sulfide with the concomitant release of acetic acid. In bacteria such as *Escherichia coli*, two forms of the enzyme are known (genes *cysK* and *cysM*). In plants there are also two forms, one located in the cytoplasm and the other in chloroplasts. Cystathionine beta-synthase [2] catalyzes the first irreversible step in homocysteine transsulfuration; the conjugation of homocysteine and serine forming cystathionine. Like CSase it is a pyridoxal-phosphate dependent enzyme. The two types of enzymes are evolutionarily related. The pyridoxal-phosphate group of CSases has been shown to be attached to a lysine residue which is located in the N-terminal section of these enzymes; the sequence around this residue is highly conserved and can be used as a signature pattern to detect this class of enzymes.

Consensus pattern: K-x-E-x(3)-[PA]-[STAGC]-x-S-[IVAP]-K-x-R-x-[STAG]-x(2)-[LIVM] [The 2nd K is the pyridoxal-P attachment site]

[1] Saito K., Kurosawa M., Murakoshi I. FEBS Lett. 328:111-114(1993).[2] Swaroop M., Bradley K., Ohura T., Tahara T., Roper M.D., Rosenberg L.E., Kraus J.P. J. Biol. Chem. 267:11455-11461(1992).

602. S locus glycop

S-locus glycoprotein family. In Brassicaceae, self-incompatible plants have a self/non-self

Comment: recognition system. This is sporophytically controlled by multiple alleles at a single locus (S). S-locus glycoproteins, Comment: as well as S-receptor kinases, are in linkage with the S-alleles [1]. Number of members: 128

[1] Evolutionary aspects of the S-related genes of the Brassica self-incompatibility system: synonymous and nonsynonymous base substitutions. Hinata K, Watanabe M, Yamakawa S, Satta Y, Isogai A; Genetics 1995;140:1099-1104. [2] Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gene (SLR1) in *Raphanus sativus* L. and self-incompatible ornamental plants in the Brassicaceae. Sakamoto K, Kusaba M, Nishio T; Mol Gen Genet 1998;258:397-403.

603. (sdh cyt) Succinate dehydrogenase cytochrome b subunit signatures

Succinate dehydrogenase (SDH) is a membrane-bound complex of two main components: a membrane-extrinsic component composed of an FAD-binding flavoprotein and an iron-sulfur protein, and a hydrophobic component composed of a cytochrome B and a membrane anchor protein. The cytochrome b component is a mono heme transmembrane protein [1,2,3] belonging to a family that groups: - Cytochrome b-556 from bacterial SDH (gene *sdhC*). - Cytochrome b560 from the mammalian mitochondrial SDH complex. - Cytochrome b560 subunit encoded in the mitochondrial genome of some algae and in the plant *Marchantia polymorpha*. - Cytochrome b from yeast mitochondrial SDH complex (gene *SDH3* or *CYB3*). - Protein *cyt-1* from *Caenorhabditis*. These cytochromes are proteins of about 130 residues that comprise three transmembrane regions. There are two conserved histidines which may be involved in binding the heme group. Two signature patterns have been developed that include these histidine residues.

Consensus pattern: R-P-[LIVMT]-x(3)-[LIVM]-x(6)-[LIVMWPK]-x(4)-S-x(2)-H-R-x-[ST] [H could be a heme ligand]

Consensus pattern: H-x(3)-[GA]-[LIVMT]-R-[HF]-[LIVMF]-x-[FYWM]-D-x-[GVA] [H could be a heme ligand]

[1] Yu L., Wei Y.-Y., Usui S., Yu C.-A. J. Biol. Chem. 267:24508-24515(1992). [2] Abraham P.R., Mulder A., Van't Riet J., Raue H.A. Mol. Gen. Genet. 242:708-716(1994). [3] Leblanc C., Boyen C., Richard O., Bonnard G., Grienemberger J.M., Kloareg B. J. Mol. Biol. 250:484-495(1995).

604. Sec1 family

[1] The Sec1 family: a novel family of proteins involved in synaptic transmission and general secretion. Halachmi N, Lev Z; J Neurochem 1996;66:889-897.

Number of members: 40

605. Protein secE/sec61-gamma signature

In bacteria, the secE protein plays a role in protein export; it is one of the components - with secY and secA - of the preprotein translocase. In eukaryotes, the evolutionary related protein sec61-gamma plays a role in protein translocation through the endoplasmic reticulum; it is part of a trimeric complex that also consists of sec61-alpha and beta [1]. Both secE and sec61-gamma are small proteins of about 60 to 90 amino acids that contain a single transmembrane region at their C-terminal extremity (Escherichia coli secE is an exception, in that it possesses an extra N-terminal segment of 60 residues that contains two additional transmembrane domains). The sequence of secE/sec61-gamma is not extremely well conserved, however it is possible to derive a signature pattern centered on a conserved proline located 10 residues before the beginning of the transmembrane domain.

Consensus pattern: [LIVMFY]-x(2)-[DENQGA]-x(4)-[LIVMFTA]-x-[KRV]-x(2)-[KW]-P-x(3)-[SEQ]-x(7)-[LIVT]-[LIVGA]-[LIVFGAST]

[1] Hartmann E., Sommer T., Prehn S., Goerlich D., Jentsch S., Rapoport T.A. Nature 367:654-657(1994).

606. 11-S plant seed storage proteins signature

Plant seed storage proteins, whose principal function appears to be the major nitrogen source for the developing plant, can be classified, on the basis of their structure, into different families. 11-S are non-glycosylated proteins which form hexameric structures [1,2]. Each of the subunits in the hexamer is itself composed of an acidic and a basic chain derived from a single precursor and linked by a disulfide bond. This structure is shown in the following representation. +-----+ ||

xxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxNGxCxxxxxxxxxxxxxxxxxxxxxx ***** <--
 ---Acidic-subunit----->-----Basic-subunit-----> <-----About 480-to-500-
 residues----->'C': conserved cysteine involved in a disulfide bond.'*': position of the
 pattern. Proteins that belong to the 11-S family are: pea and broad bean legumins, rape

cruciferin, rice glutelins, cotton beta-globulins, soybean glycinins, pumpkin 11-S globulin, oat globulin, sunflower helianthinin G3, etc. The region that includes the conserved cleavage site between the acidic and basic subunits (Asn-Gly) and a proximal cysteine residue which is involved in the interchain disulfide bond have been used as a signature pattern for this family of proteins.

Consensus pattern: N-G-x-[DE](2)-x-[LIVMF]-C-[ST]-x(11,12)-[PAG]-D [C is involved in a disulfide bond]

[1] Hayashi M., Mori H., Nishimura M., Akazawa T., Hara-Nishimura I. Eur. J. Biochem. 172:627-632(1988).[2] Shotwell M.A., Afonso C., Davies E., Chesnut R.S., Larkins B.A. Plant Physiol. 87:698-704(1988).

607. 7S seed storage protein

7S globulin is one of the main storage proteins of most angiosperms and gymnosperms. The 7S storage proteins are homotrimers.
Number of members: 67

[1] The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). Ko TP, Ng JD, McPherson A; Plant Physiol 1993;101:729-744.

608. Aspartate-semialdehyde dehydrogenase signature

Aspartate-semialdehyde dehydrogenase (ASD) catalyzes the second step in the common biosynthetic pathway leading from Asp to diaminopimelate and Lys, to Met, and to Thr; the NADP-dependent reductive dephosphorylation of L-aspartyl phosphate to L-aspartate-semialdehyde. In bacteria and fungi, ASD is a protein of about 40 Kd (340 to 370 residues) whose sequence is not extremely well conserved [1]. A conserved cysteine residue has been implicated as important for the catalytic activity [2]. The region of conservation around the active site residue is too small to be used as signature pattern. Another more conserved region, located in the last third of the sequence, and which contains both a conserved cysteine as well as an histidine has been used instead.

Consensus pattern: [LIVM]-[SADN]-x(2)-C-x-R-[LIVM]-x(4)-[GSC]-H-[STA]

[1] Baril C., Richaud C., Fourni E., Baranton G., Saint Girons I. J. Gen. Microbiol. 138:47-53(1992).[2] Karsten W.E., Viola R.E. Biochim. Biophys. Acta 1121:234-238(1992).

N-acetyl-gamma-glutamyl-phosphate reductase active site

N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) [1,2] is the enzyme that catalyzes the third step in the biosynthesis of arginine from glutamate, the NADP-dependent reduction of N-acetyl-5-glutamyl phosphate into N-acetylglutamate 5-semialdehyde. In bacteria it is a monofunctional protein of 35 to 38 Kd (gene *argC*) while in fungi it is part of a bifunctional mitochondrial enzyme (gene *ARG5,6, arg11 or arg-6*) which contains a N-terminal acetylglutamate kinase (EC 2.7.2.8) domain and a C-terminal AGPR domain. In the *Escherichia coli* enzyme, a cysteine has been shown to be implicated in the catalytic activity, the region around this residue is well conserved and can be used as a signature pattern.

Consensus pattern: [LIVM]-[GSA]-x-P-G-C-[FY]-[AVP]-T-[GA]-x(3)-[GTAC]-[LIVM]-x-P [C is the active site residue]

[1] Ludovice M., Martin J.F., Carrachas P., Liras P. J. *Bacteriol.* 174:4606-4613(1992).[2] Gessert S.F., Kim J.H., Nargang F.E., Weiss R.L. *J. Biol. Chem.* 269:8189-8203(1994).

609. Sialyltransferase family,

Number of members: 18

610. SpoU rRNA Methylase family

This family of proteins probably use S-AdoMet. Number of members: 58

[1] SpoU protein of *Escherichia coli* belongs to a new family of putative rRNA methylases.

Koonin EV, Rudd KE; *Nucleic Acids Res* 1993;21:5519-5519. [2] The *spoU* gene of *Escherichia coli*, the fourth gene of the *spoT* operon, is essential for tRNA (Gm18) 2' methyltransferase activity. Persson BC, Jager G, Gustafsson C; *Nucleic Acids Res* 1997;25:4093-4097.

611. Stathmin family signatures

Stathmin [1] (from the Greek 'stathmos' which means relay), is an ubiquitous intracellular protein, present in a variety of phosphorylated forms and which serves as a relay for diverse

second messenger pathways. Its expression and phosphorylation are regulated throughout development and in response to extracellular signals regulating cell proliferation, differentiation and function. Stathmin is a highly conserved protein of 149 amino acid residues. Structurally, it consists of an N-terminal domain of about 45 residues followed by a

78 residue alpha-helical domain consisting of a heptad repeat coiled coil structure and a C-terminal domain of 25 residues. Protein SCG10 is a neuron-specific, membrane-associated protein that accumulates in the growth cones of developing neurons. It is highly similar in its sequence to stathmin, but differs in that it contains an additional N-terminal hydrophobic segment of 32 residues which is probably responsible for its interaction with membranes.

Xenopus protein XB3 is also evolutionary related to stathmin and also contains an additional N-terminal hydrophobic domain [2]. A conserved decapeptide which ends with the first three residues of the coiled coil domain and a second pattern that corresponds to part of the central region of the coiled coil have been selected as signatures for proteins of the stathmin family. Consensus pattern: P-[KRQ]-[KR](2)-[DE]-x-S-L-[EG]-E-

Consensus pattern: A-E-K-R-E-H-E-[KR]-E-

[1] Sobel A. Trends Biochem. Sci. 16:301-305(1991). [2] Maucuer A., Moreau J., Mechali M., Sobel A. J. Biol. Chem. 268:16420-16429(1993).

612. SUA5/yciO/yrnC family signature. The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast protein SUA5. - Escherichia coli hypothetical protein yciO and HI1198, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yrdC and HI0656, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yw1C. - Mycobacterium leprae hypothetical protein in rfe-hemK intergenic region. - Methanococcus jannaschii hypothetical protein MJ0062. These are proteins of from 20 to 46 Kd which contain a number of conserved regions in their N-terminal section. They can be picked up in the database by the following pattern.

Consensus pattern: [LIVMTA](3)-[LIVMFYC]-[PG]-T-[DE]-[STA]-x-[FY]-[GA]- [LIVM]-[GS]-

[1] Bairoch A., Rudd K.E., Robison K. Unpublished observations (1995).

613. Sucrose synthase

Sucrose synthases catalyse the synthesis of sucrose from UDP-glucose and fructose. This family includes the bulk of the sucrose synthase protein. However the carboxyl terminal region of the sucrose synthases belongs to the glycosyl transferase family Glycos. transf. 1.

614. Sulfotransferase proteins

Number of members: 59

615. Synaptophysin / synaptoporin signature

Synaptophysin and synaptoporin [1] are structurally related proteins, found in the membrane of synaptic vesicles, which may function as ionic or solute channels. These two glycoproteins seem to span the membrane four times. Both their N- and C-termini sequences seem to be cytoplasmically located. As a signature pattern for this family of proteins, a highly conserved region located in the beginning of the first intravesicular loop just after the first transmembrane domain has been selected. This region contains a cysteine residue that may be involved in a disulfide bond.

Consensus pattern: L-S-V-[DE]-C-x-N-K-T [C may be involved in a disulfide bond
[1] Knaus P., Marqueze-Pouey B., Scherer H., Betz H. Neuron 5:453-462(1990).

616. Syndecans signature

Syndecans [1,2] (from the greek syndein; to bind together) are a family of transmembrane heparan sulfate proteoglycans which are implicated in the binding of extracellular matrix components and growth factors. Syndecans bind a variety of molecules via their heparan sulfate chains and can act as receptors or as co-receptors. Structurally, these proteins consist of four separate domains: a) A signal sequence; b) An extracellular domain (ectodomain) of variable length and whose sequence is not evolutionary conserved in the various forms of syndecans. The ectodomain contains the sites of attachment of the heparan sulfate glycosaminoglycan side chains; c) A transmembrane region; d) A highly conserved

cytoplasmic domain of about 30 to 35 residues which could interact with cytoskeletal proteins. The proteins known to belong to this family are: - Syndecan 1. - Syndecan 2 or fibroglycan. - Syndecan 3 or neuroglycan or N-syndecan. - Syndecan 4 or amphiglycan or ryudocan. - Drosophila syndecan. - Caenorhabditis elegans probable syndecan (F57C7.3).The

signature pattern that has been developed for syndecans starts with the last residue of the transmembrane region and includes the first 10 residues of the cytoplasmic domain. This region, which contains four basic residues, could act as a stop transfer site.

Consensus pattern: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y

[1] Bernfield M., Kokenyesi R., Kato M., Hinkes M.T., Spring J., Gallo R.L., Lose E.J.

Annu. Rev. Cell Biol. 8:365-393(1992).[2] David G. FASEB J. 7:1023-1030(1993).

617. Syntaxin / epimorphin family signature

The following proteins have been shown to be evolutionary related [1,2,3]: - Epimorphin (or syntaxin 2), a mammalian mesenchymal protein which plays an essential role in epithelial morphogenesis. - Syntaxin 1A (also known as antigen HPC-1) and syntaxin 1B which are synaptic proteins which may be involved in docking of synaptic vesicles at presynaptic active zones. - Syntaxin 3. - Syntaxin 4, which is potentially involved in docking of synaptic vesicles at presynaptic active zones. - Syntaxin 5, which mediates endoplasmic reticulum to golgi transport. - Syntaxin 6, which is involved in intracellular vesicle trafficking. - Syntaxin 7. - Yeast PEP12 (or VPS6) which is required for the transport of proteases to the vacuole. - Yeast SED5 which is required for the fusion of transport vesicles with the Golgi complex. - Yeast SSO1 and SSO2 which are required for vesicle fusion with the plasma membrane. - Yeast VAM3, which is required for vacuolar assembly. - Arabidopsis thaliana protein KNOLLE which may be involved in cytokinesis. - Caenorhabditis elegans hypothetical proteins F35C8.4, F48F7.2, F55A11.2 and T01B11.3. The above proteins share the following characteristics: a size ranging from 30 Kd to 40 Kd; a C-terminal extremity which is highly hydrophobic and is probably involved in anchoring the protein to the membrane; a central, well conserved region, which seems to be in a coiled-coil conformation. The pattern specific for this family is based on the most conserved region of the coiled coil domain.

Consensus pattern: [RQ]-x(3)-[LIVMA]-x(2)-[LIVM]-[ESH]-x(2)-[LIVMT]-x-[DEVLM]-[LIVM]-x(2)-[LIVM]-[FS]-x(2)-[LIVM]-x(3)-[LIVT]-x(2)-Q- [GADEQ]-x(2)-[LIVM]-[DNQT]-x-[LIVMF]-[DESV]-x(2)-[LIVM]

[1] Bennett M.K., Garcia-Ararras J.E., Elferink L.A., Peterson K., Fleming A.M., Hazuka C.D., Scheiler R.H. Cell 74:863-873(1993).[2] Spring J., Kato M., Bernfield M. Trends Biochem. Sci. 18:124-125(1993).[3] Pelham H.R.B. Cell 73:425-426(1993).

5

618. Sm protein

The U1, U2, U4/U6, and U5 small nuclear ribonucleoprotein particles (snRNPs) involved in pre-mRNA splicing contain seven Sm proteins (B/B', D1, D2, D3, E, F and G) in common, which assemble around the Sm site present in four of the major spliceosomal small nuclear RNAs. These proteins contain a common sequence motif in two segments, Sm1 and Sm2, separated by a short variable linker.

10

15

[1] Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Luhrmann R EMBO J 1995;14:2076-2088. [2] Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Luhrmann R, Li J, Nagai K; Cell 1999;96:375-387.

20

619. Skp1 family

[1] Stebbins CE, Kaelin WG Jr, Pavletich NP; Science 1999;284:455-461.

25

620. Protein secY signatures

The eubacterial secY protein [1] plays an important role in protein export. It interacts with the signal sequences of secretory proteins as well as with two other components of the protein translocation system: secA and secE. SecY is an integral plasma membrane protein of 419 to 492 amino acid residues that apparently contains ten transmembrane segments. Such a structure probably confers to secY a 'translocator' function, providing a channel for periplasmic and outer-membrane precursor proteins. Homologs of secY are found in archaeobacteria [2]. SecY is also encoded in the chloroplast genome of some algae [3] where it could be involved in a prokaryotic-like protein export system across the two membranes of

30

the chloroplast endoplasmic reticulum (CER) which is present in chromophyte and cryptophyte algae. Two signature patterns have been developed for secY proteins. The first corresponds to the second transmembrane region, which is the most conserved section of these proteins. The second spans the C-terminal part of the fourth transmembrane region, a short intracellular loop, and the N-terminal part of the fifth transmembrane region.

Consensus pattern: [GST]-[LIVMF](2)-x-[LIVM]-G-[LIVM]-x-P-[LIVMFY](2)-x-[AS]-[GSTQ]-[LIVMFAT](3)-Q-[LIVMFA](2)

Consensus pattern: [LIVMFYW](2)-x-[DE]-x-[LIVMF]-[STN]-x(2)-G-[LIVMF]-[GST]-[NST]-G-x-[GST]-[LIVMF](3)

[1] Ito K. Mol. Microbiol. 6:2423-2428(1992).[2] Auer J., Spicker G., Boeck A. Biochimie 73:683-688(1991).[3] Douglas S.E. FEBS Lett. 298:93-96(1992).

621. (Seed protein) Small hydrophilic plant seed proteins signature. The following small hydrophilic plant seed proteins are structurally related: - Arabidopsis thaliana proteins GEA1 and GEA6. - Cotton late embryogenesis abundant (LEA) protein D-19. - Carrot EMB-1 protein. - Barley LEA proteins B19.1A, B19.1B, B19.3 and B19.4. - Maize late embryogenesis abundant protein Emb564. - Radish late seed maturation protein p8B6. - Rice embryonic abundant protein Emp1. - Sunflower 10 Kd late embryogenesis abundant protein (DS10). - Wheat Em proteins. These proteins contain from 83 to 153 amino acid residues and may play a role [1,2] in equipping the seed for survival, maintaining a minimal level of hydration in the dry organism and preventing the denaturation of cytoplasmic components. They may also play a role during imbibition by controlling water uptake. As a signature pattern, the best conserved region in the sequence of these proteins has been developed, it is a glycine-rich nonapeptide located in the N-terminal section.

Consensus pattern: G-[EQ]-T-V-V-P-G-G-T-

[1] Dure L. III, Crouch M., Harada J., Ho T.-H. D., Mundy J., Quatrano R., Thomas T., Sung Z.R. Plant Mol. Biol. 12:475-486(1989).[2] Gaubier P., Raynal M., Hull G., Huestis G.M., Grellet F., Arenas C., Pages M., Delseny M. Mol. Gen. Genet. 238:409-418(1993).

622. Serine carboxypeptidases, active sites

All known carboxypeptidases are either metallo carboxypeptidases or serinecarboxypeptidases. The catalytic activity of the serine carboxypeptidases, like that of the trypsin family serine proteases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine [1]. Proteins known to be serine carboxypeptidases are: - Barley and wheat serine carboxypeptidases I, II, and III [2]. - Yeast carboxypeptidase Y (YSCY) (gene PRC1), a vacuolar protease involved in degrading small peptides. - Yeast KEX1 protease, involved in killer toxin and alpha-factor precursor processing. - Fission yeast *sxa2*, a probable carboxypeptidase involved in degrading or processing mating pheromones [3]. - *Penicillium janthinellum* carboxypeptidase S1 [4]. - *Aspergillus niger* carboxypeptidase pepF. - *Aspergillus satoii* carboxypeptidase cpdS. - Vertebrate protective protein / cathepsin A [5], a lysosomal protein which is not only a carboxypeptidase but also essential for the activity of both beta-galactosidase and neuraminidase. - Mosquito vitellogenic carboxypeptidase (VCP) [6]. - *Naegleria fowleri* virulence-related protein Nf314 [7]. - Yeast hypothetical protein YBR139w. - *Caenorhabditis elegans* hypothetical proteins C08H9.1, F13D12.6, F32A5.3, F41C3.5 and K10B2.2. This family also includes: - Sorghum (s)-hydroxymandelonitrile lyase (hydroxynitrile lyase) (HNL) [8], an enzyme involved in plant cyanogenesis. The sequences surrounding the active site serine and histidine residues are highly conserved in all these serine carboxypeptidases.

Consensus pattern: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS] [S is the active site residue]

Consensus pattern: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x-[IVAQ]-P-x(3)-[PSA] [H is the active site residue]

[1] Liao D.L., Remington S.J. J. Biol. Chem. 265:6528-6531(1990).[2] Sorensen S.B., Svendsen I., Breddam K. Carlsberg Res. Commun. 54:193-202(1989).[3] Imai Y., Yamamoto M. Mol. Cell. Biol. 12:1827-1834(1992).[4] Svendsen I., Hofmann T., Endrizzi J., Remington J., Breddam K. FEBS Lett. 333:39-43(1993).[5] Galjart N.J., Morreau H., Willemsen R., Gillemans N., Bonten E.J., d'Azzo A. J. Biol. Chem. 266:14754-14762(1991).[6] Cho W.L., Deitsch K.W., Raikhel A.S. Proc. Natl. Acad. Sci. U.S.A. 88:10821-10824(1991).[7] Hu W.N., Kopachik W., Band R.N. Infect. Immun. 60:2418-2424(1992).[8] Wajant H., Mundry K.W., Pfitzenmaier K. Plant Mol. Biol. 26:735-746(1994).[9] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

623. Serpins signature. Serpins (SERine Proteinase INhibitors) [1,2,3,4] are a group of structurally related proteins. They are high molecular weight (400 to 500 amino acids), extracellular, irreversible serine protease inhibitors with a well defined structural-functional characteristic: a reactive region that acts as a 'bait' for an appropriate serine protease. This region is found in the C-terminal part of these proteins. Proteins which are known to belong to the serpin family are listed below (references are only provided for recently determined sequences): - Alpha-1 protease inhibitor (alpha-1-antitrypsin, contrapsin). - Alpha-1-antichymotrypsin, - Antithrombin III. - Alpha-2-antiplasmin. - Heparin cofactor II. - Complement C1 inhibitor. - Plasminogen activator inhibitors 1 (PAI-1) and 2 (PAI-2). - Glia derived nexin (GDN) (Protease nexin I). - Protein C inhibitor. - Rat hepatocytes SPI-1, SPI-2 and SPI-3 inhibitors. - Human squamous cell carcinoma antigen (SCCA) which may act in the modulation of the host immune response against tumor cells. - A lepidopteran protease inhibitor. - Leukocyte elastase inhibitor which, in contrast to other serpins, is an intracellular protein. - Neuroserpin [5], a neuronal inhibitor of plasminogen activators and plasmin. - Cowpox virus crmA [6], an inhibitor of the thiol protease interleukin-1B converting enzyme (ICE). CrmA is the only serpin known to inhibit a non-serine proteinase. - Some orthopoxviruses probable protease inhibitors, which may be involved in the regulation of the blood clotting cascade and/or of the complement cascade in the mammalian host. On the basis of strong sequence similarities, a number of proteins with no known inhibitory activity are said to belong to this family: - Birds ovalbumin and the related genes X and Y proteins. - Angiotensinogen; the precursor of the angiotensin active peptide. - Barley protein Z; the major endosperm albumin. - Corticosteroid binding globulin (CBG). - Thyroxine-binding globulin (TBG). - Sheep uterine milk protein (UTMP) and pig uteroferrin-associated protein (UFAP). - Hsp47, an endoplasmic reticulum heat-shock protein that binds strongly to collagen and could act as a chaperone in the collagen biosynthetic pathway [7]. - Maspin, which seems to function as a tumor suppressor [5]. - Pigment epithelium-derived factor precursor (PEDF), a protein with a strong neutrophilic activity [8]. - Ep45, an estrogen-regulated protein from *Xenopus* [9]. A signature pattern has been developed for this family of proteins, centered on a well conserved Pro-Phe sequence which is found ten to fifteen residues on the C-terminal side of the reactive bond

Consensus pattern: [LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-
[LIVMFYC]-x-[LIVMFAH]-

[1] Carrell R., Travis J. Trends Biochem. Sci. 10:20-24(1985).[2] Carrell R., Pemberton
P.A., Boswell D.R. Cold Spring Harbor Symp. Quant. Biol. 52:527-535(1987).[3] Huber R.,
Carrell R.W. Biochemistry 28:8951-8966(1989).[4] Remold-O'Donneel E. FEBS Lett.
315:105-108(1993).[5] Osterwalder T., Contartese J., Stoeckli E.T., Kuhn T.B., Sonderegger
P. EMBO J. 15:2944-2953(1996).[6] Komiyama T., Ray C.A., Pickup D.J., Howard A.D.,
Thornberry N.A., Peterson E.P., Salvesen G. J. Biol. Chem. 269:19331-19337(1994).[7]
Clarke E., Sandwal B.D. Biochim. Biophys. Acta 1129:246-248(1992).[8] Zou Z.,
Anisowicz A., Neveu M., Rafidi K., Sheng S., Sager R., Hendrix M.J., Seftor E., Thor A.
Science 263:526-529(1994).[9] Steele F.R., Chader G.J., Johnson L.V., Tombran-Tink J.
Proc. Natl. Acad. Sci. U.S.A. 90:1526-1530(1993).[10] Holland L.J., Suksang C., Wall A.A.,
Roberts L.R., Moser D.R., Bhattacharya A. J. Biol. Chem. 267:7053-7059(1992).

624. Sigma-54 interaction domain signatures and profile

Some bacterial regulatory proteins activate the expression of genes from promoters
recognized by core RNA polymerase associated with the alternative sigma-54 factor. These
have a conserved domain of about 230 residues involved in the ATP-dependent [1,2]
interaction with sigma-54. This domain has been found in the proteins listed below: - *acoR*
from *Alcaligenes eutrophus*, an activator of the acetoin catabolism operon *acoXABC*. - *algB*
from *Pseudomonas aeruginosa*, an activator of alginate biosynthetic gene *algD*. - *dctD* from
Rhizobium, an activator of *dctA*, the C4-dicarboxylate transport protein. - *dhaR* from
Citrobacter freundii, a regulator of the *dha* operon for glycerol utilization. - *fhlA* from
Escherichia coli, an activator of the formate dehydrogenase H and hydrogenase III structural
genes. - *flbD* from *Caulobacter crescentus*, an activator of flagellar genes. - *hoxA* from
Alcaligenes eutrophus, an activator of the hydrogenase operon. - *hrpS* from *Pseudomonas*
syringae, an activator of *hprD* as well as other *hrp* loci involved in plant pathogenicity. -
hupR1 from *Rhodobacter capsulatus*, an activator of the [NiFe] hydrogenase genes *hupSL*. -
hydG from *Escherichia coli* and *Salmonella typhimurium*, an activator of the hydrogenase
activity. - *levR* from *Bacillus subtilis*, which regulates the expression of the levanase operon
(*levDEFG* and *sacC*). - *nifA* (as well as *anfA* and *vnfA*) from various bacteria, an activator of

the *nif* nitrogen-fixing operon. - *ntrC*, from various bacteria, an activator of nitrogen assimilatory genes such as that for glutamine synthetase (*glnA*) or of the *nif* operon. - *pgtA* from *Salmonella typhimurium*, the activator of the inducible phospho- glycerate transport system. - *pilR* from *Pseudomonas aeruginosa*, an activator of pilin gene transcription. - *rocR* from *Bacillus subtilis*, an activator of genes for arginine utilization - *tyrR* from *Escherichia coli*, involved in the transcriptional regulation of aromatic amino-acid biosynthesis and transport. - *wtsA*, from *Erwinia stewartii*, an activator of plant pathogenicity gene *wtsB*. - *xylR* from *Pseudomonas putida*, the activator of the *tol* plasmid xylene catabolism operon *xylCAB* and of *xylS*. - *Escherichia coli* hypothetical protein *yfhA*. - *Escherichia coli* hypothetical protein *yhgB*. About half of these proteins (*algB*, *dcdT*, *flbD*, *hoxA*, *hupR1*, *hydG*, *ntrC*, *pgtA* and *pilR*) belong to signal transduction two-component systems [3] and possess a domain that can be phosphorylated by a sensor-kinase protein in their N- terminal section. Almost all of these proteins possess a helix-turn-helix DNA-binding domain in their C-terminal section. The domain which interacts with the sigma-54 factor has an ATPase activity. This may be required to promote a conformational change necessary for the interaction [4]. The domain contains an atypical ATP-binding motif A (P-loop) as well as a form of motif B. The two ATP-binding motifs are located in the N-terminal section of the domain; signature patterns have been developed for both motifs. Other regions of the domain are also conserved. One of them, located in the C-terminal section, has been selected as a third signature pattern.

Consensus pattern: [LIVMFY](3)-x-G-[DEQ]-[STE]-G-[STAV]-G-K-x(2)-[LIVMFY]

Consensus pattern: [GS]-x-[LIVMF]-x(2)-A-[DNEQASH]-[GNEK]-G-[STIM]-[LIVMFY](3)-[DE]-[EK]-[LIVM]

Consensus pattern: [FYW]-P-[GS]-N-[LIVM]-R-[EQ]-L-x-[NHAT]

[1] Morrett E., Segovia L. J. *Bacteriol.* 175:6067-6074(1993).[2] Austin S., Kundrot C., Dixon R. *Nucleic Acids Res.* 19:2281-2287(1991).[3] Albright L.M., Huala E., Ausubel F.M. *Annu. Rev. Genet.* 23:311-336(1989).[4] Austin S., Dixon R. *EMBO J.* 11:2219-2228(1992).

625. Sigma-70 factors family signatures

Sigma factors [1] are bacterial transcription initiation factors that promote the attachment of the core RNA polymerase to specific initiation sites and are then released. They alter the

specificity of promoter recognition. Most bacteria express a multiplicity of sigma factors. Two of these factors, sigma-70 (gene rpoD), generally known as the major or primary sigma factor, and sigma-54 (gene rpoN or ntrA) direct the transcription of a wide variety of genes. The other sigma factors, known as alternative sigma factors, are required for the transcription of specific subsets of genes. With regard to sequence similarity, sigma factors can be grouped into two classes: the sigma-54 and sigma-70 families. The sigma-70 family includes, in addition to the primary sigma factor, a wide variety of sigma factors, some of which are listed below: - *Bacillus* sigma factors involved in the control of sporulation-specific genes: sigma-E (sigE or spoIIGB), sigma-F (sigF or spoIIAC), sigma-G (sigG or spoIIIG), sigma-H (sigH or spo0C) and sigma-K (sigK or spoIVCB/spoIIIC). - *Escherichia coli* and related bacteria sigma-32 (gene rpoH or htpR) involved in the expression of heat shock genes. - *Escherichia coli* and related bacteria sigma-27 (gene fliA) involved in the expression of the flagellin gene. - *Escherichia coli* sigma-S (gene rpoS or katF) which seems to be involved in the expression of genes required for protection against external stresses. - *Myxococcus xanthus* sigma-B (sigB) which is essential for the late-stage differentiation of that bacteria. Alignments of the sigma-70 family permit the identification of four regions of high conservation [2,3]. Each of these four regions can in turn be subdivided into a number of sub-regions. Signature patterns based on the two best-conserved sub-regions have been developed. The first pattern corresponds to sub-region 2.2; the exact function of this sub-region is not known although it could be involved in the binding of the sigma factor to the core RNA polymerase. The second pattern corresponds to sub-region 4.2 which seems to harbor a DNA-binding 'helix-turn-helix' motif involved in binding the conserved -35 region of promoters recognized by the major sigma factors. The second pattern starts one residue before the N-terminal extremity of the HTH region and ends six residues after its C-terminal extremity.

Consensus pattern: [DE]-[LIVMF](2)-[HEQS]-x-G-x-[LIVMFA]-G-L-[LIVMFYE]-x-[GSAM]-[LIVMAP]

Consensus pattern: [STN]-x(2)-[DEQ]-[LIVM]-[GAS]-x(4)-[LIVMF]-[PSTG]-x(3)-[LIVMA]-x-[NQR]-[LIVMA]-[EQH]-x(3)-[LIVMFW]-x(2)-[LIVM]

[1] Helmann J.D., Chamberlin M.J. Annu. Rev. Biochem. 57:839-872(1988).[2] Gribskov M., Burgess R.R. Nucleic Acids Res. 14:6745-6763(1986).[3] Lonetto M.A., Gribskov M., Gross C.A. J. Bacteriol. 174:3843-3849(1992).[4] Lonetto M.A., Brown K.L., Rudd K.E., Buttner M.J. Proc. Natl. Acad. Sci. U.S.A. 91:7573-7577(1994).

626. Signal carboxyl-terminal domain. 430 members.

627. Signal peptidases I signatures

Signal peptidases (SPases) [1] (also known as leader peptidases) remove the signal peptides from secretory proteins. In prokaryotes three types of SPases are known: type I (gene *lepB*) which is responsible for the processing of the majority of exported pre-proteins; type II (gene *lsp*) which only process lipoproteins, and a third type involved in the processing of pili subunits. SPase I is an integral membrane protein that is anchored in the cytoplasmic membrane by one (in *B. subtilis*) or two (in *E. coli*) N-terminal transmembrane domains with the main part of the protein protruding in the periplasmic space. Two residues have been shown [2,3] to be essential for the catalytic activity of SPase I: a serine and an lysine. SPase I is evolutionary related to the yeast mitochondrial inner membrane protease subunit 1 and 2 (genes *IMP1* and *IMP2*) which catalyze the removal of signal peptides required for the targeting of proteins from the mitochondrial matrix, across the inner membrane, into the inter-membrane space [4]. In eukaryotes the removal of signal peptides is effected by an oligomeric enzymatic complex composed of at least five subunits: the signal peptidase complex (SPC). The SPC is located in the endoplasmic reticulum membrane. Two components of mammalian SPC, the 18 Kd (SPC18) and the 21 Kd (SPC21) subunits as well as the yeast SEC11 subunit have been shown [5] to share regions of sequence similarity with prokaryotic SPases I and yeast *IMP1/IMP2*. Three signature patterns for these proteins have been developed. The first signature contains the putative active site serine, the second signature contains the putative active site lysine which is not conserved in the SPC subunits, and the third signature corresponds to a conserved region of unknown biological significance which is located in the C-terminal section of all these proteins.

Consensus pattern: [GS]-x-S-M-x-[PS]-[AT]-[LF] [S is an active site residue]

Consensus pattern: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY] [K is an active site residue]

Consensus pattern: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG]

[1] Dalbey R.E., von Heijne G. Trends Biochem. Sci. 17:474-478(1992).[2] Sung M., Dalbey R.E. J. Biol. Chem. 267:13154-13159(1992).[3] Black M.T. J. Bacteriol. 175:4957-4961(1993).[4] Nunnari J., Fox T.D., Walter P. Science 262:1997-2004(1993).[5] van Dijk

J.M., de Jong A., Vehmaanpera J., Venema G., Bron S. EMBO J. 11:2819-2828(1992).[6]

Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

5 628. (sodcu) Copper/Zinc superoxide dismutase signatures

Copper/Zinc superoxide dismutase (SODC) [1] is one of the three forms of an enzyme that catalyzes the dismutation of superoxide radicals. SODC binds one atom each of zinc and copper. Various forms of SODC are known: acytoplasmic form in eukaryotes, an additional chloroplast form in plants, an extracellular form in some eukaryotes, and a periplasmic form in prokaryotes. The metal binding sites are conserved in all the known SODC sequences [2]. Two signature patterns have been derived for this family of enzymes: the first one contains two histidine residues that bind the copper atom; the second one is located in the C-terminal section of SODC and contains a cysteine which is involved in a disulfide bond.

Consensus pattern: [GA]-[IMFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGDE] [The two H's are copper ligands]

Consensus pattern: G-[GN]-[SGA]-G-x-R-x-[SGA]-C-x(2)-[IV] [C is involved in a disulfide bond]

[1] Bannister J.V., Bannister W.H., Rotilio G. CRC Crit. Rev. Biochem. 22:111-154(1987).[

2] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:175-184(1992).

629. (sodfe) Manganese and iron superoxide dismutases signature

Manganese superoxide dismutase (SODM) [1] is one of the three forms of an enzyme that catalyzes the dismutation of superoxide radicals. The four ligands of the manganese atom are conserved in all the known SODM sequences. These metal ligands are also conserved in the related iron form of superoxide dismutases [2,3]. A short conserved region which includes two of the four ligands: an aspartate and a histidine has been selected as a signature.

Consensus pattern: D-x-W-E-H-[STA]-[FY](2) [D and H are manganese/iron ligands]

[1] Bannister J.V., Bannister W.H., Rotilio G. CRC Crit. Rev. Biochem. 22:111-154(1987).[

2] Parker M.W., Blake C.C.F. FEBS Lett. 229:377-382(1988).[3] Smith M.W., Doolittle R.F. J. Mol. Evol. 34:175-184(1992).

630. Spectrin repeat

Spectrin repeats are found in several proteins involved in cytoskeletal structure. These include spectrin, alpha-actinin and dystrophin. The sequence repeat used in this family is taken from the structural repeat in reference [2]. The spectrin repeat forms a three helix bundle. The second helix is interrupted by proline in some sequences.

Number of members: 898

[1] Actin-binding proteins. 1: Spectrin super family. Hartwig JH; Protein Profile 1995;2:732-732. [2] Crystal structure of the repetitive segments of spectrin. Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D; Science 1993;262:2027-2030.

631. (subtilase) Streptomyces subtilisin-type inhibitors signature

Bacteria of the Streptomyces family produce a family of proteinase inhibitors [1] characterized by their strong activity toward subtilisin. They are collectively known as SSIs: Streptomyces Subtilisin Inhibitors. Some SSIs also inhibit trypsin or chymotrypsin. In their mature secreted form, SSIs are proteins of about 110 residues with two conserved disulfide bonds. +-----+ +-----+ |||
 xxxxxxxxxxxxxxxCxxxxxxxxCxxxxxxxxCx#xxxxxxxxxxxxCxxxxxxxx *****C';
 conserved cysteine involved in a disulfide bond. '#': active site residue. '*': position of the pattern.

Consensus pattern: C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L [The two C's are involved in a disulfide bond]

[1] Taguchi S., Kojima S., Terabe M., Miura K.-I., Momose H. Eur. J. Biochem. 220:911-918(1994).

632. Sugar transport proteins signatures

In mammalian cells the uptake of glucose is mediated by a family of closely related transport proteins which are called the glucose transporters [1,2,3]. At least seven of these transporters are currently known to exist (in Human they are encoded by the GLUT1 to GLUT7 genes). These integral membrane proteins are predicted to comprise twelve membrane spanning domains. The glucose transporters show sequence similarities [4,5] with a number

of other sugar or metabolite transport proteins listed below (references are only provided for recently determined sequences). - *Escherichia coli* arabinose-proton symport (araE). - *Escherichia coli* galactose-proton symport (galP). - *Escherichia coli* and *Klebsiella pneumoniae* citrate-proton symport (also known as citrate utilization determinant) (gene cit).

5 - *Escherichia coli* alpha-ketoglutarate permease (gene kgtP). - *Escherichia coli* proline/betaine transporter (gene proP) [6]. - *Escherichia coli* xylose-proton symport (xylE). - *Zymomonas mobilis* glucose facilitated diffusion protein (gene glf). - Yeast high and low affinity glucose transport proteins (genes SNF3, HXT1 to HXT14). - Yeast galactose transporter (gene GAL2). - Yeast maltose permeases (genes MAL3T and MAL6T). - Yeast

10 myo-inositol transporters (genes ITR1 and ITR2). - Yeast carboxylic acid transporter protein homolog JEN1. - Yeast inorganic phosphate transporter (gene PHO84). - *Kluyveromyces lactis* lactose permease (gene LAC12). - *Neurospora crassa* quinate transporter (gene Qa-y), and *Emmericella nidulans* quinate permease (gene qutD). - *Chlorella* hexose carrier (gene HUP1). - *Arabidopsis thaliana* glucose transporter (gene STP1). - Spinach sucrose

15 transporter. - *Leishmania donovani* transporters D1 and D2. - *Leishmania enriettii* probable transport protein (LTP). - Yeast hypothetical proteins YBR241c, YCR98c and YFL040w. - *Caenorhabditis elegans* hypothetical protein ZK637.1. - *Escherichia coli* hypothetical proteins yabE, ydjE and yhjE. - *Haemophilus influenzae* hypothetical proteins HI0281 and HI0418. - *Bacillus subtilis* hypothetical proteins yxbC and yxdF. It has been suggested [4] that these

20 transport proteins have evolved from the duplication of an ancestral protein with six transmembrane regions, this hypothesis is based on the conservation of two G-R-[KR] motifs. The first one is located between the second and third transmembrane domains and the second one between transmembrane domains 8 and 9. Two patterns have been developed to detect this family of proteins. The first pattern is based on the G-R-[KR] motif; but because this

25 motif is too short to be specific to this family of proteins, a pattern from a larger region centered on the second copy of this motif was derived. The second pattern is based on a number of conserved residues which are located at the end of the fourth transmembrane segment and in the short loop region between the fourth and fifth segments.

Consensus pattern: [LIVMSTAG]-[LIVMFSAG]-x(2)-[LIVMSA]-[DE]-x-[LIVMFYWA]-

30 G- R-[RK]-x(4,6)-[GSTA]

Consensus pattern: [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)-[RK]

[1] Silverman M. Annu. Rev. Biochem. 60:757-794(1991).[2] Gould G.W., Bell G.I. Trends Biochem. Sci. 15:18-23(1990).[3] Baldwin S.A. Biochim. Biophys. Acta 1154:17-49(1993).[

4] Maiden M.C.J., Davis E.O., Baldwin S.A., Moore D.C.M., Henderson P.J.F. *Nature* 325:641-643(1987).[5] Henderson P.J.F. *Curr. Opin. Struct. Biol.* 1:590-601(1991).[6] Culham D.E., Lasby B., Marangoni A.G., Milner J.L., Steer B.A., van Nues R.W., Wood J.M. *J. Mol. Biol.* 229:268-276(1993).

5

633. Synaptobrevin signature

Synaptobrevin [1] is an intrinsic membrane protein of small synaptic vesicles whose function is not yet known, but which is highly conserved in mammals, electric ray (where its is known as VAMP-1), *Drosophila* and yeast [2]. In yeast there are two closely related forms of synaptobrevin (genes *SNC1* and *SNC2*) while in mammals there is at least 4 (genes *SYB1*, *SYB2*, *SYB3* and *SYBL1*). Structurally synaptobrevin consist of a N-terminal cytoplasmic domain of from 90 to 110 residues, followed by a transmembrane region, and then by a short (from 2 to 22 residues) C-terminal intravesicular domain. As a signature pattern for synaptobrevin, a highly conserved stretch of residues located in the central part of the sequence was selected.

Consensus pattern: N-[LIVM]-[DENS]-[KL]-V-x-[DEQ]-R-x(2)-[KR]-[LIVM]-[STDE]- x-[LIVM]-x-[DE]-[KR]-[TA]-[DE]

[1] Suedhof T.C., Baumert M., Perin M.S., Jahn R. *Neuron* 2:1475-1481(1989).[2] Gerst J.E., Rodgers L., Riggs M., Wigler M. *Proc. Natl. Acad. Sci. U.S.A.* 89:4338-4342(1992).

Downloaded from www.worldscientific.com by UNIVERSITY OF CALIFORNIA on 06/16/16. For personal use only.

634. TBC domain. Identification of a TBC domain in GYP6_YEAST and GYP7_YEAST, which are GTPase activator proteins of yeast Ypt6 and Ypt7, imply that these domains are GTPase activator proteins of Rab-like small GTPases. Number of members: 55

[1] Medline: 96032578. Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16. Richardson PM, Zon LI; Oncogene 1995;11:1139-1148.

[2]Medline: 97398935. A shared domain between a spindle assembly checkpoint protein and Ypt/Rab-specific GTPase-activators. Neuwald AF; Trends Biochem Sci 1997;22:243-244.

635. Transcription factor TFIID repeat signature (TBP)

Transcription factor TFIID (or TATA-binding protein, TBP) [1,2] is a general factor that plays a major role in the activation of eukaryotic genes transcribed by RNA polymerase II. TFIID binds specifically to the TATA box promoter element which lies close to the position of transcription initiation. There is a remarkable degree of sequence conservation of a C-terminal domain of about 180 residues in TFIID from various eukaryotic sources. This region is necessary and sufficient for TATA box binding. The most significant structural feature of this domain is the presence of two conserved repeats of a 77 amino-acid region. The intramolecular symmetry generates a saddle-shaped structure that sits astride the DNA [3]. Drosophila TRF (TBP-related factor) [4] is a sequence-specific transcription factor that also binds to the TATA box and is highly similar to TFIID. Archaeobacteria also possess a TBP homolog [5]. A signature pattern that spans the last 50 residues of the repeated region has been derived.-

Consensus pattern: Y-x-P-x(2)-[IF]-x(2)-[LIVM](2)-x-[KRH]-x(3)-P-[RKQ]-x(3)-L-[LIVM]-F-x-[STN]-G-[KR]-[LIVM]-x(3)-G-[TAGL]-[KR]-x(7)-[AGC]-x(7)-[LIVM] [1] Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M., Roeder R.G. Nature 346:387-390(1990).[2] Gash A., Hoffmann A., Horikoshi M., Roeder R.G., Chua N.-H. Nature 346:390-394(1990).[3] Nikolov D.B., Hu S.-H., Lin J., Gasch A., Hoffmann A., Horikoshi M., Chua N.-H., Roeder R.G., Burley S.K. Nature 360:40-46(1992).[4] Crowley T.E., Hoey T., Liu J.-K., Jan Y.N., Jan L.Y., Tjian R. Nature 361:557-561(1993).[5] Marsh T.L., Reich C.I., Whitelock R.B., Olsen G.J. Proc. Natl. Acad. Sci. U.S.A. 91:4180-4184(1994).

636. Translationally controlled tumor protein signatures (TCTP)

Mammalian translationally controlled tumor protein (TCTP) (or P23) is a protein which has been found to be preferentially synthesized in cells during the early growth phase of some types of tumor [1,2], but which is also expressed in normal cells. The physiological function of TCTP is still not known. It is a hydrophilic protein of 18 to 20 Kd. Close homologs have been found in plants [3], earthworm [4], *Caenorhabditis elegans* (F52H2.11), *Hydra*, budding yeast (YKL056c) [5] and fission yeast (SpAC1F12.02c). Two of the best conserved regions have been selected as signature patterns for TCTP.

Consensus pattern: [IFA]-[GA]-[GAS]-N-[PAK]-S-[GA]-E-[GDE]-[PAGE]-[DEQGA]

Consensus pattern: [FLVH]-[FY]-[IVCT]-G-E-x-[MA]-x(2,5)-[DEN]-[GAST]-x-[LV]-[AV]-x(3)-[FYW]

[1] Boehm H., Beendorf R., Gaestel M., Gross B., Nuernberg P., Kraft R., Otto A., Bielka H. *Biochem. Int.* 19:277-286(1989). [2] Makrides S., Chitpatima S.T., Bandyopadhyay R., Brawerman G. *Nucleic Acids Res.* 16:2350-2350(1988). [3] Pay A., Heberle-Bors E., Hirt H. *Plant Mol. Biol.* 19:501-503(1992). [4] Stuerzenbaum S.R., Kille P., Morgan A.J. *Biochim. Biophys. Acta* 1398:294-304(1998). [5] Rasmussen S.W. *Yeast* 10:S63-S68(1994).

637. TFIIS zinc ribbon domain signature

Transcription factor S-II (TFIIS) [1] is a eukaryotic protein necessary for efficient RNA polymerase II transcription elongation, past template-encoded pause sites. TFIIS shows DNA-binding activity only in the presence of RNA polymerase II. It is a protein of about 300 amino acids whose sequence is highly conserved in mammals, *Drosophila*, yeast (where it was first known as PPR2, a transcriptional regulator of *URA4*, and then as *DST1*, the DNA strand transfer protein alpha [2]) and in the archaeobacteria *Sulfolobus acidocaldarius* [3]. This family also includes the eukaryotic and archeobacterial RNA polymerase subunits of the 15 Kd / M family (see <PDOC00790>) as well as the following viral proteins: - *Vaccinia* virus RNA polymerase 30 Kd subunit (rpo30) [4]. - African swine fever virus protein I243L [5]. The best conserved region of all these proteins contains four cysteines that bind a zinc ion and fold in a conformation termed a 'zinc ribbon' [6]. Besides these cysteines, there are a

number of other conserved residues which can be used to help define a specific pattern for this type of domain.

Consensus pattern: C-x(2)-C-x(9)-[LIVMQSAR]-[QH]-[STQL]-[RA]-[SACR]-x-[DE]-[DET]-[PGSEA]-x(6)-C-x(2,5)-C-x(3)-[FW] [The four C's are zinc ligands]

- 5 [1] Hirashima S., Hirai H., Nakanishi Y., Natori S. J. Biol. Chem. 263:3858-3863(1988).[2]
Kipling D., Kearsey S.E. Nature 353:509-509(1991).[3] Langer D., Zillig W. Nucleic Acids
Res. 21:2251-2251(1993).[4] Ahn B.-Y., Gershon P.D., Jones E.V., Moss B. Mol. Cell. Biol.
10:5433-5441(1990).[5] Rodriguez J.M., Salas M.L., Vinuela E. Virology 186:40-52(1992).[
6] Qian X., Jeon C., Yoon H., Agarwal K., Weiss M.A. Nature 365:277-279(1993).

638. Tetrahydrofolate dehydrogenase/cyclohydrolase signatures (THF DHG CYH)

Enzymes that participate in the transfer of one-carbon units are involved in various biosynthetic pathways. In many of these processes the transfers of one-carbon units are mediated by the coenzyme tetrahydrofolate (THF). Various reactions generate one-carbon derivatives of THF which can be interconverted between different oxidation states by formyltetrahydrofolate synthetase (EC 6.3.4.3), methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5 or EC 1.5.1.15) and methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9). The dehydrogenase and cyclohydrolase activities are expressed by a variety of multifunctional enzymes: - Eukaryotic C-1-tetrahydrofolate synthase (C1-THF synthase), which catalyzes all three reactions described above. Two forms of C1-THF synthases are known [1], one is located in the mitochondrial matrix, while the second one is cytoplasmic. In both forms the dehydrogenase/cyclohydrolase domain is located in the N-terminal section of the 900 amino acids protein and consists of about 300 amino acid residues. The C1-THF synthases are NADP- dependent. - Eukaryotic mitochondrial bifunctional dehydrogenase/cyclohydrolase [2]. This is an homodimeric NAD-dependent enzyme of about 300 amino acid residues. - Bacterial fold [3]. Fold is an homodimeric bifunctional NADP-dependent enzyme of about 290 amino acid residues. The sequence of the dehydrogenase/cyclohydrolase domain is highly conserved in all forms of the enzyme. Two conserved regions have been selected as signature patterns. The first one is located in the N-terminal part of these enzymes and contains three acidic residues. The second pattern is a highly conserved sequence of 9 amino acids which is located in the C-terminal section.

Consensus pattern: [EQ]-x-[EQK]-[LIVM](2)-x(2)-[LIVM]-x(2)-[LIVMY]-N-x-[DN]-x(5)-[LIVMF](3)-Q-L-P-[LV]

Consensus pattern: P-G-G-V-G-P-[MF]-T-[IV]

[1] Shannon K.W., Rabinowitz J.C. J. Biol. Chem. 263:7717-7725(1988).[2] Belanger C.,

- 5 Mackenzie R.E. J. Biol. Chem. 264:4837-4843(1989).[3] d'Ari L., Rabinowitz J.C. J. Biol. Chem. 266:23953-23958(1991).

639. Triosephosphate isomerase active site (TIM)

- 10 Triosephosphate isomerase (EC 5.3.1.1) (TIM) [1] is the glycolytic enzyme that catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. TIM plays an important role in several metabolic pathways and is essential for efficient energy production. It is a dimer of identical subunits, each of which is made up of about 250 amino-acid residues. A glutamic acid residue is involved in the catalytic mechanism [2]. The
- 15 sequence around the active site residue is perfectly conserved in all known TIM's and can be used as a signature pattern for this type of enzyme.

Consensus pattern: [AV]-Y-E-P-[LIVM]-W-[SA]-I-G-T-[GK] [E is the active site residue]

[1] Lolis E., Alber T., Davenport R.C., Rose D., Hartman F.C., Petsko G.A. Biochemistry 29:6609-6618(1990).[2] Knowles J.R. Nature 350:121-124(1991).

640. Thymidine kinase cellular-type signature (TK)

Thymidine kinase (TK) (EC 2.7.1.21) is an ubiquitous enzyme that catalyzes the ATP-dependent phosphorylation of thymidine. A comparison of TK sequences has shown [1,2,3]

- 25 that there are two different families of TK. One family groups together TK from herpes viruses as well as cellular thymidylate kinases, while the second family currently consists of TK from the following sources: - Vertebrates. - Bacterial. - Bacteriophage T4. - Pox viruses. - African swine fever virus (ASF). - Fish lymphocystis disease virus (FLDV). A conserved region which is located in the C-terminal section of these enzymes has been selected as a
- 30 signature pattern for this family of TKA.

Consensus pattern: [GA]-x(1,2)-[DE]-x-Y-x-[STAP]-x-C-[NKR]-x-[CH]-[LIVMFYWH]

[1] Boyle D.B., Coupar B.E.H., Gibbs A.J., Seigman L.J., Both G.W. Virology 156:355-365(1987).[2] Blasco R., Lopez-Otin C., Munoz M., Bockamp E.-O., Simon-Mateo C.,

Vinuela E. *Virology* 178:301-304(1990).[3] Robertson G.R., Whalley J.M. *Nucleic Acids Res.* 16:11303-11317(1988).

641. Thymidine kinase from herpesvirus (TK herpes)

[1]

Medline: 96003730

Crystal structures of the thymidine kinase from herpes simplex virus type-1 in complex with deoxythymidine and ganciclovir.

Brown DG, Visse R, Sandhu G, Davies A, Rizkallah PJ, Melitz

C, Summers WC, Sanderson MR;

Nat Struct Biol 1995;2:876-881.

Number of members: 65

642. Nuclear transition protein 2 signatures (TP2)

In mammals, the second stage of spermatogenesis is characterized by the conversion of nucleosomal chromatin to the compact, non-nucleosomal and transcriptionally inactive form found in the sperm nucleus. This condensation is associated with a double-protein transition. The first transition corresponds to the replacement of histones by several spermatid-specific proteins, also called transition proteins, which are themselves replaced by protamines during the second transition. Nuclear transition protein 2 (TP2) is one of those spermatid-specific proteins. TP2 is a basic, zinc-binding protein [1] of 116 to 137 amino-acid residues.

Structurally, TP2 consists of three distinct parts: a conserved serine-rich N-terminal domain of about 25 residues, a variable central domain of 20 to 50 residues which contains cysteine residues, and a conserved C-terminal domain of about 70 residues rich in lysines and arginines. Two signature patterns for TP2 have been developed: one located in the N-terminal domain, the other in the C-terminal.

Consensus pattern: H-x(3)-H-S-[NS]-S-x-P-Q-S

Consensus pattern: K-x-R-K-x(2)-E-G-K-x(2)-K-[KR]-K

[1] Baskaran R., Rao M.R.S. *Biochem. Biophys. Res. Commun.* 179:1491-1499(1991).

643. Thiamine pyrophosphate enzymes signature (TTP enzymes)

A number of enzymes require thiamine pyrophosphate (TPP) (vitamin B1) as a cofactor. It has been shown [1] that some of these enzymes are structurally related. These related TPP enzymes are: - Pyruvate oxidase (POX) (EC 1.2.3.3) Reaction catalyzed: pyruvate + orthophosphate + O(2) + H(2)O = acetyl phosphate + CO(2) + H(2)O(2). - Pyruvate decarboxylase (PDC) (EC 4.1.1.1) Reaction catalyzed: pyruvate = acetaldehyde + CO(2). - Indolepyruvate decarboxylase (EC 4.1.1.74) [2] Reaction catalyzed: indole-3-pyruvate = indole-3-acetaldehyde + CO(2). - Acetolactate synthase (ALS) (EC 4.1.3.18) Reaction catalyzed: 2 pyruvate = acetolactate + CO(2). - Benzoylformate decarboxylase (BFD) (EC 4.1.1.7) [3] Reaction catalyzed: benzoylformate = benzaldehyde + CO(2). A conserved region which is located in their C-terminal section has been selected as a signature pattern for these enzymes.

Consensus pattern: [LIVMF]-[GSA]-x(5)-P-x(4)-[LIVMFYW]-x-[LIVMF]-x-G-D-[GSA]-[GSAC]

[1] Green J.B.A. FEBS Lett. 246:1-5(1989). [2] Koga J., Adachi T., Hidaka H. Mol. Gen. Genet. 226:10-16(1991). [3] Tsou A.Y., Ransom S.C., Gerlt J.A., Buechter D.D., Babbitt P.C., Kenyon G.L. Biochemistry 29:9856-9862(1990).

644. TPR Domain

[1]

Medline: 95397415

Tetratricopeptide repeat interactions: to TPR or not to TPR?

Lamb JR, Tugendreich S, Hieter P;

Trends Biochem Sci 1995;20:257-259.

[2]Medline: 98151343

The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.

Das AK, Cohen PW, Barford D;

EMBO J 1998;17:1192-1199.

Number of members: 621

645. Uroporphyrin-III C-methyltransferase signatures (TP methylase)

Uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (SUMT) [1,2] catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to the C-2 and C-7atoms of uroporphyrinogen III to yield precorrin-2 via the intermediate formation of precorrin-1. SUMT is the first enzyme specific to the cobalamin pathway and precorrin-2 is a common intermediate in the biosynthesis of corrinoids such as vitamin B12, siroheme and coenzyme F430. The sequences of SUMT from a variety of eubacterial and archaebacterial species are currently available. In species such as *Bacillus megaterium* (gene *cobA*), *Pseudomonas denitrificans* (*cobA*) or *Methanobacterium ivanovii* (gene *corA*) SUMT is a protein of about 25 to 30 Kd. In *Escherichia coli* and related bacteria, the *cysG* protein, which is involved in the biosynthesis of siroheme, is a multifunctional protein composed of a N-terminal domain, probably involved in transforming precorrin-2 into siroheme, and a C-terminal domain which has SUMT activity. The sequence of SUMT is related to that of a number of *P. denitrificans* and *Salmonella typhimurium* enzymes involved in the biosynthesis of cobalamin which also seem to be SAM-dependent methyltransferases [3,4]. The similarity is especially strong with two of these enzymes: *cobI/cbiL* which encodes S-adenosyl-L-methionine--precorrin-2 methyltransferase and *cobM/cbiF* whose exact function is not known. Two signature patterns have been developed for these enzymes. The first corresponds to a well conserved region in the N-terminal extremity (called region 1 in [1,3]) and the second to a less conserved region located in the central part of these proteins (this pattern spans what are called regions 2 and 3 in [1,3]).

Consensus pattern: [LIVM]-[GS]-[STAL]-G-P-G-x(3)-[LIVMFY]-[LIVM]-T-[LIVM]-[KRHQG]-[AG]

Consensus pattern: V-x(2)-[LI]-x(2)-G-D-x(3)-[FYW]-[GS]-x(8)-[LIVF]-x(5,6)-[LIVMFYWPAC]-x-[LIVMY]-x-P-G

[1] Blanche F., Robin C., Couder M., Faucher D., Cauchois L., Cameron B., Crouzet J. J. *Bacteriol.* 173:4637-4645(1991).[2] Robin C., Blanche F., Cauchois L., Cameron B., Couder M., Crouzet J. J. *Bacteriol.* 173:4893-4896(1991).[3] Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C., Blanche F., Thibaut D., Debussche L. J. *Bacteriol.* 172:5980-5990(1990).[4] Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S., Church G.M. J.

Bacteriol. 175:3303-3316(1993).[5] Mattheakis L.C., Shen W.H., Collier R.J. Mol. Cell. Biol. 12:4026-4037(1992).

5 646. Tudor domain

Domain of unknown function present in several RNA-binding proteins. copies in the Drosophila Tudor protein. Slight ambiguities in the alignment. Number of members: 18 [1]Medline: 97200561 Tudor domains in proteins that interact with RNA. Ponting CP; Trends Biochem Sci 1997;22:51-52. [2]Medline: 97157029 The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in Drosophila melanogaster development. Callebaut I, Mornon JP; Biochem J 1997;321:125-132.

15 647. Terpene synthase family

It has been suggested that this gene family be designated tps (for terpene synthase) [1]. It has been split into six subgroups on the basis of phylogeny, called tpsa-tpsf. tpsa includes vetispiradiene synthase Swiss:Q39979, 5-epi-aristolochene synthase, Swiss:Q40577 and (+)-delta-cadinene synthase Swiss:P93665. tpsb includes (-)-limonene synthase, Swiss:Q40322. tpsc includes kaurene synthase A, Swiss:O04408. tpsd includes taxadiene synthase, Swiss:Q41594, pinene synthase, Swiss:O24475 and myrcene synthase, Swiss:O24474. tpse includes kaurene synthase B. tpsf includes linalool synthase.

Number of members: 51

[1]

30 Medline: 97413772

Monoterpene synthases from grand fir (*Abies grandis*). cDNA isolation, characterization, and functional expression of myrcene synthase, (-)-(4S)-limonene synthase, and

(-)-(1S,5S)-pinene synthase.

Bohlmann J, Steele CL, Croteau R;
J Biol Chem 1997;272:21784-21792.

5

648. ThiF family

This family contains a repeated domain in ubiquitin
activating enzyme E1 and members of the bacterial
ThiF/MoeB/HesA family. Number of members: 87

10

649. Thioester dehydrase

Members of this family are involved in fatty acid biosynthesis.
Number of members: 19

15

[1]

Medline: 96398612

Structure of a dehydratase-isomerase from the bacterial
pathway for biosynthesis of unsaturated fatty acids: two
catalytic activities in one active site.

20

Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL;
Structure 1996;4:253-264.

Database Reference: SCOP; 1mka; fa; [SCOP-USA][CATH-PDBSUM]

Database reference: PFAMB; PB058036;

25

650. Tub family signatures

The mouse tubby mutation is the cause of maturity-onset obesity, insulin resistance and
sensory deficits. This mutation maps to a gene, tub [1,2], which codes for a protein that
belongs to a family which currently consists of the following members: - Mammalian tub, an
hydrophilic protein of about 500 residues, which could be involved in the hypothalamic
regulation of body weight. - Human protein TULP1 [3] which may be involved in retinitis
pigmentosa 14, a retinal degeneration disease. - Mouse protein p4-6 whose function is not
known. - Caenorhabditis elegans hypothetical protein F10B5.4. - Several fragmentary

30

sequences from plants, Drosophila and human ESTs. While the N-terminal part of these protein is not conserved in length nor in the sequence, the C-terminal 250 residues are highly conserved. Therefore, two regions were selected in the C-terminal part as signature patterns. The second region is located at the C-terminal extremity and contains a penultimate cysteine residue that could be critical to the normal functioning of these proteins.

Consensus pattern: F-[KHQ]-G-R-V-[ST]-x-A-S-V-K-N-F-Q

Consensus pattern: A-F-[AG]-I-[SAC]-[LIVM]-[ST]-S-F-x-[GST]-K-x-A-C-E

[1] Kleyn P.W., Fan W., Kovats S.G., Lee J.L., Pulido J.C., Wu Y., Berkemeier L.R., Misumi D.J., Holmgren L., Charlat O., Woolf E.A., Tayber O., Brody T., Shu P., Hawkins F., Kennedy B., Baldini L., Ebeling C., Alperin G.D., Deeds J., Lakey N.D., Culpepper J., Chen H., Gluecksmann-Kuis M.A., Carlson G.A., Duyk G.M., Moore K.J. *Cell* 85:281-290(1996).[2] Noben-Trauth K., Naggert J.K., North M.A., Nishina P.M. *Nature* 380:534-538(1996).[3] North M.A., Naggert J.K., Yan Y., Noben-Trauth K., Nishina P.M. *Proc. Natl. Acad. Sci. U.S.A.* 94:3128-3133(1997).

651. Eukaryotic DNA topoisomerase I active site

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type I topoisomerases act by catalyzing the transient breakage of DNA, one strand at a time, and the subsequent rejoining of the strands. When a eukaryotic type I topoisomerase breaks a DNA backbone bond, it simultaneously forms a protein-DNA link where the hydroxyl group of a tyrosine residue is joined to a 3'-phosphate on DNA, at one end of the enzyme-severed DNA strand. In eukaryotes and pox virus topoisomerases I, there are a number of conserved residues in the region around the active site tyrosine.

Consensus pattern: [DEN]-x(6)-[GS]-[IT]-S-K-x(2)-Y-[LIVM]-x(3)-[LIVM] [Y is the active site tyrosine]

[1] Sternglanz R. *Curr. Opin. Cell Biol.* 1:533-535(1990).[2] Sharma A., Mondragon A. *Curr. Opin. Struct. Biol.* 5:39-47(1995).[3] Lynn R.M., Bjornsti M.-A., Caron P.R., Wang J.C. *Proc. Natl. Acad. Sci. U.S.A.* 86:3559-3563(1989).[4] Roca J. *Trends Biochem. Sci.* 20:156-160(1995).[E1]

652. Transaldolase signatures

Transaldolase (EC 2.2.1.2) catalyzes the reversible transfer of a three-carbonketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form erythrose 4-phosphate and fructose 6-phosphate. This enzyme, together with transketolase, provides a link between the glycolytic and pentose-phosphate pathways. Transaldolase is an enzyme of about 34 Kd whose sequence has been well conserved throughout evolution. A lysine has been implicated [1] in the catalytic mechanism of the enzyme; it acts as a nucleophilic group that attacks the carbonyl group of fructose-6-phosphate. Transaldolase is evolutionary related [2] to a bacterial protein of about 20Kd (known as talC in *Escherichia coli*), whose exact function is not yet known. Two signature patterns have been developed for these proteins. The first, located in the N-terminal section, contains a perfectly conserved pentapeptide; these cond. includes the active site lysine.

Consensus pattern: [DG]-[IVSA]-T-[ST]-N-P-[STA]-[LIVMF](2)

Consensus pattern: [LIVM]-x-[LIVM]-K-[LIVM]-[PAS]-x-[ST]-x-[DENQPAS]-G- [LIVM]-x-[AGV]-x-[QEKIRST]-x-[LIVM] [K is the active site residue]

[1] Miosga T., Schaaff-Gerstenschlaeger I., Franken E., Zimmermann F.K. *Yeast* 9:1241-1249(1993). [2] Reizer J., Reizer A., Saier M.H. Jr. *Microbiology* 141:961-971(1995).

653. (Transpeptidase) Penicillin binding protein transpeptidase domain

The active site serine (residue 337 in [Swiss:P14677](#)) is conserved in all members of this family.

[1] Pares S, Mouz N, Petillot Y, Hakenbeck R, Dideberg O *Nat Struct Biol* 1996;3:284-289.

654. Trehalase signatures

Trehalase (EC 3.2.1.28) is the enzyme responsible for the degradation of the disaccharide alpha, alpha-trehalose yielding two glucose subunits [1]. It is an enzyme found in a wide variety of organisms and whose sequence has been highly conserved throughout evolution. Two of the most highly conserved regions have been selected as signature patterns. The first pattern is located in the central section, the second one is in the C-terminal region.

Consensus pattern: P-G-G-R-F-x-E-x-Y-x-W-D-x-Y

Consensus pattern: Q-W-D-x-P-x-[GA]-W-[PAS]-P

[1] Kopp M., Mueller H., Holzer H. J. Biol. Chem. 268:4766-4774(1993).[2] Henrissat B.,

Bairoch A. Biochem. J. 293:781-788(1993).[E1]

5

655. Trehalose-6-phosphate synthase domain

OtsA (Trehalose-6-phosphate synthase) is homologous to regions in the subunits of yeast trehalose-6-phosphate synthase/phosphate complex, [1].

[1] Kaasen I, McDougall J, Strom AR; Gene 1994;145:9-15.

10

656. Tropomyosins signature

Tropomyosins [1,2] are family of closely related proteins present in muscle and non-muscle cells. In striated muscle, tropomyosin mediate the interactions between the troponin complex and actin so as to regulate muscle contraction. The role of tropomyosin in smooth muscle and non-muscle tissues is not clear. Tropomyosin is an alpha-helical protein that forms a coiled-coil dimer. Muscle isoforms of tropomyosin are characterized by having 284 amino acid residues and a highly conserved N-terminal region, whereas non-muscle forms are generally smaller and are heterogeneous in their N-terminal region. The signature pattern for tropomyosins is based on a very conserved region in the C-terminal section of tropomyosins and which is present in both muscle and non-muscle forms.

Consensus pattern: L-K-E-A-E-x-R-A-E

[1] Smilie L.B. Trends Biochem. Sci. 4:151-155(1979).[2] McLeod A.R. BioEssays 6:208-212(1986).

25

657. Troponin

Troponin (Tn) contains three subunits, Ca²⁺ binding (TnC), inhibitory (TnI), and tropomyosin binding (TnT). this Pfam contains members of the TnT subunit.

30

Troponin is a complex of three proteins, Ca²⁺ binding (TnC), inhibitory (TnI), and tropomyosin binding (TnT).

The troponin complex regulates Ca^{++} induced muscle contraction.

This family includes troponin T and troponin I. Troponin I binds to actin and troponin T binds to tropomyosin.

Number of members: 81 [1]

5 Medline: 87144593

Structure of co-crystals of tropomyosin and troponin.

White SP, Cohen C, Phillips GN Jr;

Nature 1987;325:826-828. [2]

Medline: 95155315

10 A direct regulatory role for troponin T and a dual role for troponin C in the Ca^{2+} regulation of muscle contraction.

Potter JD, Sheng Z, Pan BS, Zhao J;

J Biol Chem 1995;270:2557-2562.

[3]Medline: 95324796

15 The troponin complex and regulation of muscle contraction.

Farah CS, Reinach FC;

FASEB J 1995;9:755-767.

20 658. (Tryp mucin) Mucin-like glycoprotein

This family of trypanosomal proteins resemble vertebrate mucins. The protein consists of three regions. The N and C termini are conserved between all members of the family, whereas the central region is not well conserved and contains a large number of threonine residues which can be glycosylated [1].

25

Indirect evidence suggested that these genes might encode the core protein of parasite mucins, glycoproteins that were proposed to be involved in the interaction with, and invasion of, mammalian host cells.

30 [1] Di Noia JM, Sanchez DO, Frasch AC; J Biol Chem 1995;270:24146-24149.

[2] Di Noia JM, D'Orso I, Aslund L, Sanchez DO, Frasch AC; J Biol Chem 1998;273:10843-10850.

659. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. A few years ago it was found [2] that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiary structure based on a Rossmann fold. Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMF]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

660. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1b)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. A few years ago it was found [2]

that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiary structure based on a Rossmann fold. Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

661. (tRNA-synt 1C) tRNA synthetases class I (E and Q)

Other tRNA synthetase sub-families are too dissimilar to be included.

This family includes only glutamyl and glutaminyl tRNA synthetases.

In some organisms, a single glutamyl-tRNA synthetase aminoacylates both tRNA(Glu) and tRNA(Gln).

[1] Rath VL, Silvian LF, Beijer B, Sproat BS, Steitz TA; Structure 1998;6:439-449.

662. (tRNA-synt 1d) tRNA synthetases class I (R)

Other tRNA synthetase sub-families are too dissimilar to be included.

This family includes only arginyl tRNA synthetase.

663. Aminoacyl-transfer RNA synthetases class-II signatures (tRNA synt 2)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]

Consensus pattern: [GSTALVF]-[DENQHRKP]-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Delarue M., Moras D. BioEssays 15:675-687(1993).[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991). [5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).[6] Cusack S. Biochimie 75:1077-1081(1993).[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

664. Aminoacyl-transfer RNA synthetases class-I signature (tRNA synt 1e)

Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. A few years ago it was found [2]

that several aminoacyl-tRNA synthetases share a region of similarity in their N-terminal section, in particular the consensus tetrapeptide His-Ile-Gly-His ('HIGH') is very well conserved. The 'HIGH' region has been shown [3] to be part of the adenylate binding site. The 'HIGH' signature has been found in the aminoacyl-tRNA synthetases specific forarginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. These aminoacyl-tRNA synthetases are referred to as class-I synthetases [4,5,6] and seem to share the same tertiary structure based on a Rossmann fold. Consensus pattern: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMF]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC]

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Webster T., Tsai H., Kula M., Mackie G.A., Schimmel P. Science 226:1315-1317(1984).[3] Brick P., Bhat T.N., Blow D.M. J. Mol. Biol. 208:83-98(1988).[4] Delarue M., Moras D. BioEssays 15:675-687(1993).[5] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[6] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).

665. Aminoacyl-transfer RNA synthetases class-II signatures (tRNA synt 2b)
Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]
Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

Thaumatin [1] is an intensively sweet-tasting protein (100 000 times sweeter than sucrose on a molar basis) from *Thaumatococcus daniellii*, an African brush. The protein is made of about 200 residues and contains 8 disulfide bonds. A number of proteins have been found to be related to thaumatin. These proteins are listed below (references are only provided for recently determined sequences). - A maize alpha-amylase/trypsin inhibitor. - Two tobacco pathogenesis-related proteins: PR-R major and minor forms, which are induced after infection with viruses. - Salt-induced protein NP24 from tomato. - Osmotin, a salt-induced protein from tobacco. - Osmotin-like proteins OSML13, OSML15 and OSML81 from potato [2]. - P21, a leaf protein from soybean. - PWIR2, a leaf protein from wheat. - Zeamatin, a maize antifungal protein [3]. The exact biological function of all these proteins is not yet known. A conserved region that includes three cysteine residues known (in thaumatin) to be involved in disulfide bonds has been selected as a signature pattern.

A diagram showing a horizontal line with a dashed segment in the middle, flanked by solid segments. The dashed segment is labeled with a plus sign and a vertical bar. The solid segments are labeled with a plus sign and a vertical bar, and a vertical bar and asterisks respectively.

xxCXXXXXXXXXXXXXXXXXCxxCxxCxCXXXXXXXXXXXXXXXXXCxxCxCxxxCxCxxCCxCxxxCxxxxxC
xxxCx ||||| |||| | +-+ +-+ | +---+ +---+ +---+ | +-----+ 'C': conserved cysteine
involved in a disulfide bond.^{†*}: position of the pattern.

Consensus pattern: G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-G-x(2,3)-C

[1] Edens L., Heslinga L., Klok R., Ledeboer A.M., Maat J., Toonen M.Y., Visser C.,

Verriss C.T. Gene 18:1-12(1982).[2] Zhu B., Chen T.H.H., Li P.H. Plant Physiol. 108:929-937(1995).[3] Malchorn D.E., Borgmeyer J.R., Smith C.E., Shah D.M.; Plant Physiol. 106:1471-1481(1994).

667. Thiolases signatures

Two different types of thiolase [1,2,3] are found both in eukaryotes and in prokaryotes: acetoacetyl-CoA thiolase (EC 2.3.1.9) and 3-ketoacyl-CoA thiolase (EC 2.3.1.16). 3-ketoacyl-CoA thiolase (also called thiolase I) has a broad chain-length specificity for its substrates and is involved in degradative pathways such as fatty acid beta-oxidation. Acetoacetyl-CoA thiolase (also called thiolase II) is specific for the thiolysis of acetoacetyl-CoA and involved in biosynthetic pathways such as poly beta-hydroxybutyrate synthesis or steroid biogenesis. In eukaryotes, there are two forms of 3-ketoacyl-CoA thiolase: one located in the mitochondrion and the other in peroxisomes. There are two conserved cysteine residues important for thiolase activity. The first located in the N-terminal section of the enzymes is involved in the formation of an acyl-enzyme intermediate; the second located at the C-terminal extremity is the active site base involved in deprotonation in the condensation reaction. Mammalian nonspecific lipid-transfer protein (nsL-TP) (also known as sterol carrier protein 2) is a protein which seems to exist in two different forms: a 14 Kd protein (SCP-2) and a larger 58 Kd protein (SCP-x). The former is found in the cytoplasm or the mitochondria and is involved in lipid transport; the latter is found in peroxisomes. The C-terminal part of SCP-x is identical to SCP-2 while the N-terminal portion is evolutionary related to thiolases[4]. Three signature patterns have been developed for this family of proteins, two of which are based on the regions around the biologically important cysteines. The third is based on a highly conserved region in the C-terminal part of these proteins.

Consensus pattern: [LIVM]-[NST]-x(2)-C-[SAGLI]-[ST]-[SAG]-[LIVMFYNS]-x- [STAG]-[LIVM]-x(6)-[LIVM] [C is involved in formation of acyl-enzyme intermediate]

Consensus pattern: N-x(2)-G-G-x-[LIVM]-[SA]-x-G-H-P-x-[GA]-x-[ST]-G

Consensus pattern: [AG]-[LIVMA]-[STAGCLIVM]-[STAG]-[LIVMA]-C-x-[AG]-x-[AG]-x-[AG]-x-[SAG] [C is the active site residue]

[1] Peoples O.P., Sinskey A.J. J. Biol. Chem. 264:15293-15297(1989). [2] Yang S.-Y., Yang X.-Y.H., Healy-Louie G., Schulz H., Elzinga M. J. Biol. Chem. 265:10424-10429(1990). [3] Igual J.C., Gonzalez-Bosch C., Dopazo J., Perez-Ortin J.E. J. Mol. Evol. 35:147-155(1992). [4] Baker M.E., Billheimer J.T., Strauss J.F. III DNA Cell Biol. 10:695-698(1991).

668. Thioredoxin family active site

Thioredoxins [1 to 4] are small proteins of approximately one hundred amino-acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is wellconserved. Bacteriophage T4 also encodes for a thioredoxin but its primary structure is not homologous to bacterial, plant and vertebrate thioredoxins. A number of eukaryotic proteins contain domains evolutionary related tothioredoxin, all of them seem to be protein disulphide isomerases (PDI). PDI(EC 5.3.4.1) [5,6,7] is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: - PDI major isozyme; a multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deiodinase (EC 3.8. 1.4), as glutathione-insulin transhydrogenase (EC 1.8.4.2) and as a thyroid hormone-binding protein ! - ERp60 (ER-60; 58 Kd microsomal protein). ERp60 was originally thought to be a phosphoinositide-specific phospholipase C isozyme and later to be a protease. - ERp72. - P5.All PDI contains two or three (ERp72) copies of the thioredoxin domain. Bacterial proteins that act as thiol:disulfide interchange proteins thatallows disulfide bond formation in some periplasmic proteins also contain a thioredoxin domain. These proteins are: - Escherichia coli dsbA (or prfA) and its orthologs in Vibrio cholerae (tcpG) and Haemophilus influenzae (por). - Escherichia coli dsbC (or xpRA) and its orthologs in Erwinia chrysanthemi and Haemophilus influenzae. - Escherichia coli dsbD (or dipZ) and its Haemophilus influenzae ortholog. - Escherichia coli dsbE (or ccmG) and orthologs in Haemophilus influenzae, Rhodobacter capsulatus (helX), Rhizobiaceae (cycY and tlpA).

Consensus pattern: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGNTN]-C- [GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT] [The two C's form the redox-active bond]

[1] Holmgren A. Annu. Rev. Biochem. 54:237-271(1985).[2] Gleason F.K., Holmgren A. FEMS Microbiol. Rev. 54:271-297(1988).[3] Holmgren A. J. Biol. Chem. 264:13963-13966(1989).[4] Eklund H., Gleason F.K., Holmgren A. Proteins 11:13-28(1991).[5] Freedman R.B., Hawkins H.C., Murrant S.J., Reid L. Biochem. Soc. Trans. 16:96-99(1988).[6] Kivirikko K.I., Myllyla R., Pihlajaniemi T. FASEB J. 3:1609-1617(1989).[7] Freedman R.B., Hirst T.R., Tuite M.F. Trends Biochem. Sci. 19:331-336(1994).

669. (Transcript fac2) Transcription factor TFIIB repeat signature

In eukaryotes the initiation of transcription of protein encoding genes by polymerase II is modulated by general and specific transcription factors. The general transcription factors operate through common promoters elements (such as the TATA box). At least seven different proteins associates to form the general transcription factors: TFIIA, -IIB, -IID, -IIE, -IIF, -IIG, and -IIH[1]. Transcription factor IIB (TFIIB) plays a central role in the transcription of class II genes, it associates with a complex of TFIID-IIA bound to DNA (DA complex) to form a ternary complex TFIID-IIA-IBB (DAB complex) which is then recognized by RNA polymerase II [2,3]. TFIIB is a protein of about 315 to 340 amino acid residues which contains, in its C-terminal part an imperfect repeat of a domain of about 75 residues. This repeat could contribute an element of symmetry to the folded protein. The following proteins have been shown to be evolutionary related to TFIIB: - An Archaeobacterial TFIIB homolog. In *Pyrococcus woesei* a previously undetected open reading frame has been shown [4] to be highly related to TFIIB. - Fungal transcription factor IIBB 70 Kd subunit (gene PCF4/TDS4/BRF1) [5]. This protein is a general activator of RNA polymerase III transcription and plays a role analogous to that of TFIIB in pol III transcription. The central section of the repeated domain, which is the most conserved part of that domain has been selected as a signature pattern.

Consensus pattern: G-[KR]-x(3)-[STAGN]-x-[LIVMYA]-[GSTA](2)-[CSAV]-[LIVM]-[LIVMFY]-[LIVMA]-[GSA]-[STAC]

[1] Weinmann R. *Gene Expr.* 2:81-91(1992).[2] Hawley D. *Trends Biochem. Sci.* 16:317-318(1991).[3] Ha I., Lane W.S., Reinberg D. *Nature* 352:689-695(1991).[4] Ouzounis C., Sander C. *Cell* 71:189-190(1992).[5] Khoo B., Brophy B., Jackson S.P. *Genes Dev.* 8:2879-2890(1994).

670. (transcript fact) MADS-box domain signature and profile

A number of transcription factors contain a conserved domain of 56 amino-acid residues, sometimes known as the MADS-box domain [E1]. They are listed below: - Serum response factor (SRF) [1], a mammalian transcription factor that binds to the Serum Response Element (SRE). This is a short sequence of dyad symmetry located 300 bp to the 5' end of the

transcription initiation site of genes such as c-fos. - Mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). These proteins are transcription factor which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. - Drosophila myocyte-specific enhancer factor 2 (MEF2). - Yeast GRM/PRTF protein (gene MCM1) [2], a transcriptional regulator of mating-type-specific genes. - Yeast arginine metabolism regulation protein I (gene ARGR1 or ARG80). - Yeast transcription factor RLM1. - Yeast transcription factor SMP1. - Arabidopsis thaliana agamous protein (AG) [3], a probable transcription factor involved in regulating genes that determines stamen and carpel development in wild-type flowers. Mutations in the AG gene result in the replacement of the stamens by petals and the carpels by a new flower. - Arabidopsis thaliana homeotic proteins Apetala1 (AP1), Apetala3 (AP3) and Pistillata (PI) which act locally to specify the identity of the floral meristem and to determine sepal and petal development [4]. - Antirrhinum majus and tobacco homeotic protein deficiens (DEFA) and globosa (GLO) [5]. Both proteins are transcription factors involved in the genetic control of flower development. Mutations in DEFA or GLO cause the transformation of petals into sepals and of stamens into carpels. - Arabidopsis thaliana putative transcription factors AGL1 to AGL6 [6]. - Antirrhinum majus morphogenetic protein DEF H33 (squamosa). In SRF, the conserved domain has been shown [1] to be involved in DNA-binding and dimerization. A pattern that spans the complete length of the domain has been derived. The profile also spans the length of the MADS-box.

Consensus pattern: R-x-[RK]-x(5)-I-x-[DNGSK]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)-x(2)-[LIVM]-x-K(2)-A-x-E-[LIVM]-[STA]-x-L-x(4)-[LIVM]-x-[LIVM](3)-x(6)-[LIVMF]-x(2)-[FY]

[1] Norman C., Runswick M., Pollock R., Treisman R. *Cell* 55:989-1003(1988).[2] Passmore S., Maine G.T., Elble R., Christ C., Tye B.-K. *J. Mol. Biol.* 204:593-606(1988).[3] Yanofsky M., Ma H., Bowman J., Drews G., Feldmann K.A., Meyerowitz E.M. *Nature* 346:35-39(1990).[4] Goto K., Meyerowitz E.M. *Genes Dev.* 8:1548-1560(1994).[5] Troebner W., Ramirez L., Motte P., Hue I., Huijser P., Loennig W.-E., Saedler H., Sommer H., Schwartz-Sommer Z. *EMBO J.* 11:4693-4704(1992).[6] Ma H., Yanofsky M.F., Meyerowitz E.M. *Genes Dev.* 5:484-495(1991).[E1]

Transketolase (EC 2.2.1.1) (TK) catalyzes the reversible transfer of a two-carbon ketol unit from xylulose 5-phosphate to an aldose receptor, such as ribose 5-phosphate, to form sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate. This enzyme, together with transaldolase, provides a link between the glycolytic and pentose-phosphate pathways. TK requires thiamin pyrophosphate as a cofactor. In most sources where TK has been purified, it is a homodimer of approximately 70 Kd subunits. TK sequences from a variety of eukaryotic and prokaryotic sources [1,2] show that the enzyme has been evolutionarily conserved. In the peroxisomes of methylotrophic yeast *Hansenula polymorpha*, there is a highly related enzyme, dihydroxy-acetone synthase (DHAS) (EC 2.2.1.3) (also known as formaldehyde transketolase), which exhibits a very unusual specificity by including formaldehyde amongst its substrates. 1-deoxyxylulose-5-phosphate synthase (DXP synthase) [3] is an enzyme so far found in bacteria (gene *dxs*) and plants (gene *CLA1*) which catalyzes the thiamin pyrophosphate-dependent acyloin condensation reaction between carbon atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D- xylulose-5-phosphate (dxp), a precursor in the biosynthetic pathway to isoprenoids, thiamin (vitamin B1), and pyridoxol (vitamin B6). DXP synthase is evolutionary related to TK. Two regions of TK have been selected as signature patterns. The first, located in the N-terminal section, contains a histidine residue which appears to function in proton transfer during catalysis [4]. The second, located in the central section, contains conserved acidic residues that are part of the active cleft and may participate in substrate-binding [4].

Consensus pattern: R-x(3)-[LIVMTA]-[DENQSTHKF]-x(5,6)-[GSN]-G-H-[PLIVMF]-[GSTA]-x(2)-[LIMC]-[GS]

Consensus pattern: G-[DEQGSA]-[DN]-G-[PAEQ]-[ST]-[HQ]-x-[PAGM]-[LIVMYAC]-[DEFYW]-x(2)-[STAP]-x(2)-[RGA]

[1] Abedinia M., Layfield R., Jones S.M., Nixon P.F., Mattick J.S. *Biochem. Biophys. Res. Commun.* 183:1159-1166(1992).[2] Fletcher T.S., Kwee I.L., Nakada T., Largman C., Martin B.M. *Biochemistry* 31:1892-1896(1992).[3] Sprenger G.A., Schorken U., Wiegert T., Grolle S., De Graaf A.A., Taylor S.V., Begley T.P., Bringer-Meyer S., Sahm H. *Proc. Natl. Acad. Sci. U.S.A.* 94:12857-12862(1997). [4] Lindqvist Y., Schneider G., Ermiler U., Sundstroem M. *EMBO J.* 11:2373-2379(1992).

Recently a number of eukaryotic cell surface antigens have been found to be evolutionary related [1,2,3]. The proteins known to belong to this family are listed below: - Mammalian antigen CD9 (MIC3); A protein involved in platelet activation and aggregation. - Mammalian leukocyte antigen CD37, expressed on B lymphocytes. - Mammalian leukocyte antigen CD53 (OX-44), which may be involved in growth regulation in hematopoietic cells. - Mammalian lysosomal membrane protein CD63 (melanoma-associated antigen ME491; antigen AD1). - Mammalian antigen CD81 (cell surface protein TAPA-1), which may play an important role in the regulation of lymphoma cell growth. - Mammalian antigen CD82 (protein R2; antigen C33; Kangai 1 (KAI1)), which associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway. - Mammalian antigen CD151 (SFA-1; platelet-endothelial tetraspan antigen 3 (PETA-3)). - Mammalian cell surface glycoprotein A15 (TALLA-1; MXS1). - Mammalian novel antigen 2 (NAG-2). - Human tumor-associated antigen CO-029. - *Schistosoma mansoni* and *japonicum* 23 Kd surface antigen (SM23 / SJ23). These proteins share the following characteristics: they all seem to be type III membrane proteins (type III proteins are integral membrane proteins that contain a N-terminal membrane-anchoring domain which is not cleaved during biosynthesis and which functions both as a translocation signal and as a membrane anchor); they also contain three additional transmembrane regions, at least seven conserved cysteines residues, and are of approximately the same size (218 to 284 residues). These proteins are collectively known as the 'transmembrane 4 super family' (TM4) because they span the plasma membrane four times. A schematic diagram of the domain structure of these proteins is shown below.

+++++-----
 -----+-----+ || TMa | Extra | TM2| Cyt | TM3 | Extracellular | TM4 | Cyt| +-----
 +-----+---C---C---+-----CC-----C---C---+---C---+ ***** Cyt : cytoplasmic domain. TMa : transmembrane anchor. TM2 to TM4: transmembrane regions 2 to 4. 'C' : conserved cysteine. '*' : position of the pattern.

A conserved region that includes two cysteines and seems to be located in a short cytoplasmic loop between two transmembrane domains has been selected as a signature for these proteins.

Consensus pattern: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]-x(2)-[EG]-x(2)-[CWN]-[LIVM](2)

[1] Levy S., Nguyen V.Q., Andria M.L., Takahashi S. J. Biol. Chem. 266:14597-14602(1991). [2] Tomlinson M.G., Williams A.F., Wright M.D. Eur. J. Immunol. 23:136-40(1993). [3] Barclay A.N., Birkeland M.L., Brown M.H., Beyers A.D., Davis S.J., Somoza

C., Williams A.F. The leucocyte antigen factbooks. Academic Press, London / San Diego, (1993).

673. Tryptophan synthase alpha chain signature

Tryptophan synthase catalyzes the last step in the biosynthesis of tryptophan: the conversion of indoleglycerol phosphate and serine, totryptophan and glyceraldehyde 3-phosphate [1,2]. It has two functional domains: one for the aldol cleavage of indoleglycerol phosphate to indole and glyceraldehyde 3-phosphate and the other for the synthesis of tryptophan from indole and serine. In bacteria and plants [3], each domain is found on a separate subunit (alpha and beta chains), while in fungi the two domains are fused together on a single multifunctional protein. A conserved region that contains three conserved acidic residues has been selected as a signature pattern for the alpha chain. The first and the third acidic residues are believed to serve as proton donors/acceptors in the enzyme's catalytic mechanism.

Consensus pattern: [LIVM]-E-[LIVM]-G-x(2)-[FYC]-[ST]-[DE]-[PA]-[LIVMY]-[AGLI]-[DE]-G

[1] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989), [2] Hyde C.C., Miles E.W. Bio/Technology 8:27-32(1990), [3] Berlyn M.B., Last R.L., Fink G.R. Proc. Natl. Acad. Sci. U.S.A. 86:4604-4608(1989).

674. Tryptophan synthase beta chain pyridoxal-phosphate attachment site

Tryptophan synthase catalyzes the last step in the biosynthesis of tryptophan: the conversion of indoleglycerol phosphate and serine, totryptophan and glyceraldehyde 3-phosphate [1,2]. It has two functional domains: one for the aldol cleavage of indoleglycerol phosphate to indole and glyceraldehyde 3-phosphate and the other for the synthesis of tryptophan from indole and serine. In bacteria and plants [3], each domain is found on a separate subunit (alpha and beta chains), while in fungi the two domains are fused together on a single multifunctional protein. The beta chain of the enzyme requires pyridoxal-phosphate as a cofactor. The pyridoxal-phosphate group is attached to a lysine residue. The region around this lysine residue also contains two histidine residues which are part of the pyridoxal-phosphate binding site. The signature pattern for the tryptophan synthase beta chain is derived from that conserved region. -Consensus pattern: [LIVM]-x-H-x-G-[STA]-H-K-x-N [K is the pyridoxal-P attachment site]

[1] Crawford I.P. Annu. Rev. Microbiol. 43:567-600(1989).[2] Hyde C.C., Miles E.W. Bio/Technology 8:27-32(1990).[3] Berlyn M.B., Last R.L., Fink G.R. Proc. Natl. Acad. Sci. U.S.A. 86:4604-4608(1989).

5

675. Serine proteases, trypsin family, active sites

The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases [1]. A partial list of proteases known to belong to the trypsin family is shown below. - Acrosin. - Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C. - Cathepsin G. - Chymotrypsins. - Complement components C1r, C1s, C2, and complement factors B, D and I. - Complement-activating component of RA-reactive factor. - Cytotoxic cell proteases (granzymes A to H). - Duodenase I. - Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin). - Enterokinase (EC 3.4.21.9) (enteropeptidase). - Hepatocyte growth factor activator. - Hepsin. - Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin). - Plasma kallikrein. - Mast cell proteases (MCP) 1 (chymase) to 8. - Myeloblastin (proteinase 3) (Wegener's autoantigen). - Plasminogen activators (urokinase-type, and tissue-type). - Trypsins I, II, III, and IV. - Trypsases. - Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator. - Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab. - Apolipoprotein(a). - Blood fluke cercarial protease. - Drosophila trypsin like proteases: alpha, easter, snake-locus. - Drosophila protease stubble (gene sb). - Major mite fecal allergen Der p III. All the above proteins belong to family S1 in the classification of peptidases[2,E1] and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases are listed below. - Achromobacter lyticus protease I. - Lysobacter alpha-lytic protease. - Streptogrisin A and B (Streptomyces proteases A and B). - Streptomyces griseus glutamyl endopeptidase II. - Streptomyces fradiae proteases 1 and 2. Consensus pattern: [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue]

10

15

20

25

30

Consensus pattern: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-
[LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue]

[1] Brenner S. Nature 334:528-530(1988).[2] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).[E1]

5

676. (tsp) Thrombospondin type 1 domain

[1] Bork P; FEBS lett 1993;327:125-130.

10

677. Tubulin subunits alpha, beta, and gamma signature

Tubulins [1,2], the major constituent of microtubules are dimeric proteins which consist of two closely related subunits (alpha and beta). Tubulin binds two molecules of GTP at two different sites (N and E). At the E (Exchangeable) site, GTP is hydrolyzed during incorporation into the microtubule. Near the E site is an invariant region rich in glycines which is found in both chains and which is now [3] said to control the access of the nucleotide to its binding site. A signature pattern was developed from this region. With the exception of the simple eukaryotes, most species express a variety of closely related alpha and beta isotypes. In most species there is a third member of the tubulin family: gamma tubulin. Gamma tubulin is found at microtubule organizing centers (MTOC) such as the spindle poles or the centrosome, suggesting that it is involved in the minus-end nucleation of microtubule assembly [4].

Consensus pattern: [SAG]-G-G-T-G-[SA]-G

[1] Cleveland D.W., Sullivan K.F. Annu. Rev. Biochem. 54:331-365(1985).[2] Joshi H.C., Cleveland D.W. Cell Motil. Cytoskeleton 16:159-163(1990).[3] Hesse J., Thierauf M., Ponstingl H. J. Biol. Chem. 262:15472-15475(1987).[4] Joshi H.C. BioEssays 15:637-643(1993).

30 Tubulin-beta mRNA autoregulation signal

The stability of beta-tubulin mRNAs are autoregulated by their own translation product [1]. Unpolymerized tubulin subunits bind directly (or activate a factor(s) which binds co-translationally) to the nascent N-terminus of beta-tubulin. This binding is transduced through

the adjacent ribosomes to activate an RNase that degrades the polysome-bound mRNA. The recognition element has been shown to be the first four amino acids of beta-tubulin: Met-Arg-Glu-Ile. Mutations to this sequence abolish the autoregulation effect (except for the replacement of Glu by Asp); transposition of this sequence to an internal region of a polypeptide also suppresses the autoregulatory effect.

Consensus pattern: <M-R-[DE]-[IL]

[1] Cleveland D.W. Trends Biochem. Sci. 13:339-343(1988).

678. (tRNA-synt 2c) Aminoacyl-transfer RNA synthetases class-II signatures. Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure. The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7]. Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns have been derived from two of these regions.

Consensus pattern: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]-

Consensus pattern: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]-

[1] Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).[2] Delarue M., Moras D.

BioEssays 15:675-687(1993).[3] Schimmel P. Trends Biochem. Sci. 16:1-3(1991).[4] Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).[5] Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).[6] Cusack S. Biochimie 75:1077-1081(1993).[7] Cusack S., Berthet-Colominas C., Haertlein M., Nassar

N., Leberman R. Nature 347:249-255(1990).[8] Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

5 679, UBA-domain

The UBA-domain (ubiquitin associated domain) is a novel sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway. The structure of the UBA domain consists of a compact three helix bundle [1]. Number of members: 84

10 [1] Structure of a human DNA repair protein UBA domain that interacts with HIV-1
Vpr. Dieckmann T, Withers-Ward ES, Jarosinski MA, Liu CF, Chen IS, Feigon J; Nat Struct
Biol 1998;5:1042-1047.

15 680. UBX domain

Domain present in ubiquitin-regulatory proteins. Present in FAF1 and Shp1p. Number of members: 19

[1] The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. Hofmann K, Bucher P; Trends Biochem Sci 1996;21:172-173.

681. (UCH) Ubiquitin carboxyl-terminal hydrolases family 1 cysteine active site
Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol
proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of
ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well
as that of ubiquitinated proteins. There are two distinct families of UCH. The first class consist
of enzymes of about 25 Kd and is currently represented by: - Mammalian isozymes L1 and
L3. - Yeast YUH1. - Drosophila Uch. One of the active site residues of class-I UCH [3] is a
cysteine. A signature pattern has been derived from the region around that residue.

30 Consensus pattern: Q-x(3)-N-[SA]-C-G-x(3)-[LIVM](2)-H-[SA]-[LIVM]-[SA] [C is the active site residue]

[1] Jentsch S., Seufert W., Hauser H.-P. *Biochim. Biophys. Acta* 1089:127-139(1991).[2] D'andrea A., Pellman D. *Crit. Rev. Biochem. Mol. Biol.* 33:337-352(1998).[3] Johnston

S.C., Larsen C.N., Cook W.J., Wilkinson K.D., Hill C.P. *EMBO J.* 16:3787-3796(1997).[4]

Rawlings N.D., Barrett A.J. *Meth. Enzymol.* 244:461-486(1994).

5 682. Ubiquitin carboxyl-terminal hydrolases family 2 signatures (UCH-1)

Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. There are two distinct families of UCH. The second class
10 consist of large proteins (800 to 2000 residues) and is currently represented by: - Yeast UBP1, UBP2, UBP3, UBP4 (or DOA4/SSV7), UBP5, UBP7, UBP9, UBP10, UBP11, UBP12, UBP13, UBP14, UBP15 and UBP16. - Human tre-2. - Human isopeptidase T. - Human isopeptidase T-3. - Mammalian Ode-1. - Mammalian Unp. - Mouse Dub-1. - Drosophila fat facets protein (gene faf). - Mammalian faf homolog. - Drosophila D-Ubp-64E. -
15 Caenorhabditis elegans hypothetical protein R10E11.3. - Caenorhabditis elegans hypothetical protein K02C4.3. These proteins only share two regions of similarity. The first region contains a conserved cysteine which is probably implicated in the catalytic mechanism. The second region contains two conserved histidines residues, one of which is also probably implicated in the catalytic mechanism. Signature patterns for both conserved regions have
20 been developed.

Consensus pattern: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-Q [C is the putative active site residue]

Consensus pattern: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y [The two H's are putative active site residues]

25 [1] Jentsch S., Seufert W., Hauser H.-P. *Biochim. Biophys. Acta* 1089:127-139(1991).[2] D'andrea A., Pellman D. *Crit. Rev. Biochem. Mol. Biol.* 33:337-352(1998).[3] Rawlings N.D., Barrett A.J. *Meth. Enzymol.* 244:461-486(1994).

30 683. Ubiquitin carboxyl-terminal hydrolases family 2 signatures (UCH-2)

Ubiquitin carboxyl-terminal hydrolases (UCH) (deubiquitinating enzymes) [1,2] are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well

as that of ubiquitinated proteins. There are two distinct families of UCH. The second class consist of largeproteins (800 to 2000 residues) and is currently represented by: - Yeast UBP1, UBP2, UBP3, UBP4 (or DOA4/SSV7), UBP5, UBP7, UBP9, UBP10, UBP11, UBP12, UBP13, UBP14, UBP15 and UBP16. - Human tre-2. - Human isopeptidase T. - Human isopeptidase T-3. - Mammalian Ode-1. - Mammalian Unp. - Mouse Dub-1. - Drosophila facets protein (gene faf). - Mammalian faf homolog. - Drosophila D-Ubp-64E. - *Caenorhabditis elegans* hypothetical protein R10E11.3. - *Caenorhabditis elegans* hypothetical protein K02C4.3. These proteins only share two regions of similarity. The first region contains a conserved cysteine which is probably implicated in the catalytic mechanism. The second region contains two conserved histidines residues, one of which is also probably implicated in the catalytic mechanism. Signature patterns for both conserved regions have been developed.

Consensus pattern: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-Q [C is the putative active site residue]

Consensus pattern: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y [The two H's are putative active site residues]

[1] Jentsch S., Seufert W., Hauser H.-P. *Biochim. Biophys. Acta* 1089:127-139(1991). [2] D'andrea A., Pellman D. *Crit. Rev. Biochem. Mol. Biol.* 33:337-352(1998). [3] Rawlings N.D., Barrett A.J. *Meth. Enzymol.* 244:461-486(1994).

684. UDP-glycosyltransferases signature

UDP glycosyltransferases (UGT) are a superfamily of enzymes that catalyzes the addition of the glycosyl group from a UTP-sugar to a small hydrophobic molecule. This family currently consist of: - Mammalian UDP-glucuronosyl transferases (UDPGT) [1,2]. A large family of membrane-bound microsomal enzymes which catalyze the transfer of glucuronic acid to a wide variety of exogenous and endogenous lipophilic substrates. These enzymes are of major importance in the detoxification and subsequent elimination of xenobiotics such as drugs and carcinogens. - A large number of putative UDPGT from *Caenorhabditis elegans*. -

Mammalian 2-hydroxyacylsphingosine 1-beta-galactosyltransferase [3] (also known as UDP-galactose-ceramide galactosyltransferase). This enzyme catalyzes the transfer of galactose to ceramide, a key enzymatic step in the biosynthesis of galactocerebrosides, which are abundant sphingolipids of the myelin membrane of the central nervous system and peripheral

nervous system. - Plants flavonol O(3)-glucosyltransferase. An enzyme [4] that catalyzes the transfer of glucose from UDP-glucose to a flavanol. This reaction is essential and one of the last steps in anthocyanin pigment biosynthesis. - Baculoviruses ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-) [5] (egt). This enzyme catalyzes the transfer of glucose from UDP-glucose to ectysteroids which are insect molting hormones. The expression of egt in the insect host interferes with the normal insect development by blocking the molting process. - Prokaryotic zeaxanthin glucosyl transferase (gene crtX), an enzyme involved in carotenoid biosynthesis and that catalyses the glycosylation reaction which converts zeaxanthin to zeaxanthin-beta- diglucoside. - Streptomyces macrolide glucosyltransferases [6]. These enzymes specifically inactivates macrolide antibiotics via 2'-O-glycosylation using UDP-glucose. These enzymes share a conserved domain of about 50 amino acid residues located in their C-terminal section and from which a pattern has been extracted to detect them. Consensus pattern: [FW]-x(2)-O-x(2)-[LIVMYA]-[LIMV]-x(4,6)-[LVGAC]-[LVFYA]-[LIVMF]-[STAGCM]-[HNOQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]- [LIVMFA]-x(4)-[PQR]-[LIVMT]-x(3)-[PA]-x(3)-[DES]-[QEHN]

[1] Dutton G.J. (In) Glucoronidation of drugs and other compounds, Dutton G.J., Ed., pp 1-78, CRC Press, Boca Raton, (1980). [2] Burchell B., Nebert D.W., Nelson D.R., Bock K.W., Iyanagi T., Jansen P.L., Lancet D., Mulder G.J., Chowdhury J.R., Siest G., Tephly T.R., Mackenzie P.I. DNA Cell Biol. 10:487-494(1991). [3] Schulte S., Stoffel W. Proc. Natl. Acad. Sci. U.S.A. 90:10265-10269(1993). [4] Furtek D., Schiefelbein J.W., Johnston F., Nelson O.E. Jr. Plant Mol. Biol. 11:473-481(1988). [5] O'Reilly D.R., Miller L.K. Science 245:1110-1112(1989). [6] Hernandez C., Olano C., Mendez C., Salas J.A. Gene 134:139-140(1993).

685. UDP-glucose/GDP-mannose dehydrogenase family

The UDP-glucose/GDP-mannose dehydrogenases are a small group of enzymes which possesses the ability to catalyze the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate [2]. Number of members: 55

[1] Purification and characterization of guanosine diphospho-D-mannose dehydrogenase. A key enzyme in the biosynthesis of alginate by *Pseudomonas aeruginosa*. Roychoudhury S, May TB, Gill JF, Singh SK, Feingold DS, Chakrabarty AM; J Biol Chem 1989;264:9380-9385. [2] Properties and kinetic analysis of UDP-glucose dehydrogenase

from group A streptococci. Irreversible inhibition by UDP-chloroacetyl. Campbell RE, Sala RF, van de Rijn I, Tanner ME; J Biol Chem 1997;272:3416-3422.

686. Uracil-DNA glycosylase signature

Uracil-DNA glycosylase (EC 3.2.2.-) (UNG) [1] is a DNA repair enzyme that excises uracil residues from DNA by cleaving the N-glycosylic bond. Uracil in DNA can arise as a result of misincorporation of dUMP residues by DNA polymerase or deamination of cytosine. The sequence of uracil-DNA glycosylase is extremely well conserved [2] in bacteria and eukaryotes as well as in herpes viruses. More distantly related uracil-DNA glycosylases are also found in poxviruses [3]. In eukaryotic cells, UNG activity is found in both the nucleus and the mitochondria. Human UNG1 protein is transported to both the mitochondria and the nucleus [4]. The N-terminal 77 amino acids of UNG1 seem to be required for mitochondrial localization [4], but the presence of a mitochondrial transitpeptide has not been directly demonstrated. As a signature for this type of enzyme, the most N-terminal conserved region has been selected. This region contains an aspartic acid residue which has been proposed, based on X-ray structures [5,6] to act as a general base in the catalytic mechanism. Consensus pattern: [KR]-[LIV]-[LIVC]-[LIVM]-x-G-[QI]-D-P-Y [D is the active site residue]-

[1] Sancar A., Sancar G.B. Annu. Rev. Biochem. 57:29-67(1988).[2] Olsen L.C., Aasland R., Wittwer C.U., Krokan H.E., Helland D.E. EMBO J. 8:3121-3125 (1989).[3] Upton C., Stuart D.T., McFadden G. Proc. Natl. Acad. Sci. U.S.A. 90:4518-4522(1993).[4] Slupphaug G., Markussen F.-H., Olsen L.C., Aasland R., Aarsaether N., Bakke O., Krokan H.E., Helland D.E. Nucleic Acids Res. 21:2579-2584(1993).[5] Savva R., McAuley-Hecht K., Brown T., Pearl L. Nature 373:487-493(1995).[6] Mol C.D., Arvai A.S., Slupphaug G., Kavli B., Alseth I., Krokan H.E., Tainer J.A. Cell 80:869-878(1995)[7] Muller S.J., Caradonna S. Biochim. Biophys. Acta 1088:197-207(1991).[8] Meyer-Siegler K., Mauro D.J., Seal G., Wurzer J., Deriel J.K., Sirover M.A. Proc. Natl. Acad. Sci. U.S.A. 88:8460-8464(1991).[9] Muller S.J., Caradonna S. J. Biol. Chem. 268:1310-1319(1993).[10] Barnes D.E., Lindahl T., Sedgwick B. Curr. Opin. Cell Biol. 5:424-433(1993).

687. Uncharacterized protein family UPF0001 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome II hypothetical protein YBL036c. - *Caenorhabditis elegans* hypothetical protein F09E5.8. - *Bacillus subtilis* hypothetical protein ylmE. - *Escherichia coli* hypothetical protein yggS and HI0090, the corresponding *Haemophilus influenzae* protein. - *Helicobacter pylori* hypothetical protein HP0395. - *Mycobacterium tuberculosis* hypothetical protein MtCY270.20. - *Synechocystis* strain PCC 6803 hypothetical protein slr0556. - *A. Pseudomonas aeruginosa* hypothetical protein in pIT 5' region. - *A. Vibrio alginolyticus* hypothetical protein in pIT 5' region. These are proteins of from 25 to 30 Kd which contain a number of conserved regions. The best conserved region which is located in the first third of these proteins has been selected as a signature pattern.

Consensus pattern: [FW]-H-[FM]-[IV]-G-x-[LIV]-Q-x-[NKR]-K-x(3)-[LIV]

[1] Bairoch A., Rudd K.E. Unpublished observations (1996).

688. Uncharacterized protein family UPF0003 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - *Escherichia coli* protein aefA. - *Escherichia coli* hypothetical protein yggB. - *Escherichia coli* hypothetical protein yjeP and HI0195.1, the corresponding *Haemophilus influenzae* protein. - *Escherichia coli* hypothetical protein ynaI. - *Bacillus subtilis* hypothetical protein yhdY. - *Helicobacter pylori* hypothetical protein HP0415. - *Synechocystis* strain PCC 6803 hypothetical protein slr0639. - *Archaeoglobus fulgidus* hypothetical protein AF1546. - *Methanococcus jannaschii* hypothetical protein MJ0170. - *Methanococcus jannaschii* hypothetical protein MJ1143. The size of these proteins range from 30 to 120 Kd. They all contain a number of transmembrane regions. The best conserved region which is located in and just after the last potential transmembrane region has been selected as a signature pattern.

Consensus pattern: G-[STIF]-V-x(2)-[LIVM]-x(6)-[LIVMF]-x(3)-[DQ]-x(3)-[LIV]-x-[LIV]-P-N-x(2)-[LIVMF]-[LIVFSTA]-x(5)-N

[1] Bairoch A. Unpublished observations (1997).

689. Uncharacterized protein family UPF0004 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - *Escherichia coli* hypothetical protein yliG. - *Escherichia coli* hypothetical protein yleA and HI0019, the corresponding *Haemophilus influenzae* protein. - *Bacillus subtilis* hypothetical protein yqeV. - *Helicobacter pylori* hypothetical protein HP0269. - *Helicobacter pylori* hypothetical protein HP0285. - *Mycoplasma iowae* hypothetical protein in 16S RNA 5'region. - *Mycobacterium leprae* hypothetical protein B2235_C2_195. - *Pseudomonas aeruginosa* hypothetical protein in hemL 3'region. - *Synechocystis* strain PCC 6803 hypothetical protein slr0082. - *Synechocystis* strain PCC 6803 hypothetical protein slr0996. - *Methanococcus jannaschii* hypothetical protein MJ0865. - *Methanococcus jannaschii* hypothetical protein MJ0867. - *Caenorhabditis elegans* hypothetical protein F25B5.5. The size of these proteins range from 47 to 61 Kd. They contain six conserved cysteines, three of which are clustered in a region that can be used as a signature pattern.

Consensus pattern: [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVM]-x(4)-G

[1] Bairoch A. Unpublished observations (1997).

690. Uncharacterized protein family UPF0005 signature

The following proteins seem to be evolutionarily related [1]: - Mammalian protein TEGT (Testis Enhanced Gene Transcript). - *Escherichia coli* hypothetical protein yccA and HI0044, the corresponding *Haemophilus influenzae* protein. - A probable *Pseudomonas aeruginosa* ortholog of yccA. These are proteins of about 25 Kd which seem to contain seven transmembrane domains. A signature pattern that corresponds to a region that starts with the beginning of the third transmembrane domain and ends in the middle of the fourth one has been developed.

Consensus pattern: G-[LIVM](2)-[SA]-x(5,8)-G-x(2)-[LIVM]-G-P-x-L-x(4)-[SAG]-x(4,6)-[LIVM](2)-x(2)-A-x(3)-T-A-[LIVM](2)-F

[1] Walter L., Marynen P., Szpirer J., Levan G., Guenther E. *Genomics* 28:301-304(1995).

691. Uncharacterized protein family UPF0006 signatures

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome II hypothetical protein YBL055c. - *Escherichia coli* hypothetical protein

ycfH and HI0454, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yigW. - Escherichia coli hypothetical protein yjjV and HI0081, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yabD. - Haemophilus influenzae hypothetical protein HI1664. - Mycoplasma genitalium hypothetical protein MG009. These are proteins of from 24 to 47 Kd which contain a number of conserved regions. They can be picked up in the database by the following patterns.

Consensus pattern: [LIVMFY](2)-D-[STA]-H-x-H-[LIVMF]-[DN]
 Consensus pattern: P-[LIVM]-x-[LIVM]-H-x-R-x-[TA]-x-[DE]
 Consensus pattern: [LVSA]-[LIVA]-x(2)-[LIVM]-[PS]-x(3)-L-[LIVM]-[LIVMS]-E-T-D-x-

P

[1] Bairoch A., Rudd K.E. Unpublished observations (1995).

692. Uncharacterized protein family UPF0007 signature

The following proteins seems to be evolutionary related [1]: - Escherichia coli hypothetical protein ygbP and HI0672, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yacM. - Mycobacterium tuberculosis hypothetical protein MtCY06G11.29c. - Synechocystis strain PCC 6803 hypothetical protein slr0951. - A Rhodobacter capsulatus hypothetical protein in nifR3 5'region. Except for the Rhodobacter protein which contains a C-terminal extension, all these proteins have from 225 to 236 amino acids. They are hydrophilic proteins that can be picked up in the database by the following pattern.

Consensus pattern: V-L-[IV]-H-D-[GA]-A-R

[1] Bairoch A. Unpublished observations (1997).

693. Uncharacterized protein family UPF0015 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome II hypothetical protein YBR002c. - Yeast chromosome XIII hypothetical protein YMR101c. - Escherichia coli hypothetical protein yaeU and HI0920, the corresponding Haemophilus influenzae protein. - Helicobacter pylori hypothetical protein HPI221. - Mycobacterium leprae hypothetical protein B1937_F2_65. - A Corynebacterium glutamicum hypothetical protein in aroF 3'region. - A Streptomyces fradiae hypothetical

protein in transposon Tn4556. - *Synechocystis* strain PCC 6803 hypothetical protein slI0505. - *Methanococcus jannaschii* hypothetical protein MJ1372. These are proteins of about 26 to 40 Kd whose central region is well conserved. They can be picked up in the database by the following pattern.

- 5 Consensus pattern: [DE]-[LIVMF](3)-R-T-[SG]-G-x(2)-R-x-S-x-[FY]-[LIVM](2)-W-Q-
[1] Wolfe K.H., Lohan A.J.E. *Yeast* 10:S41-S46(1994).

694. Uncharacterized protein family UPF0016 signature

- 10 The following uncharacterized proteins have been shown [1] to share regions of similarities: -
Yeast hypothetical protein YBR187w. - Fission yeast hypothetical protein SpAC17G8.08c. -
Mouse protein pFT27. - *Synechocystis* strain PCC 6803 hypothetical protein slI0615. These
are hydrophobic proteins of 200 to 320 amino acids that seem to contain six or seven
transmembrane domains. A conserved region which seems, in the eukaryotic proteins of this
15 family, to directly follow the second transmembrane domain has been selected as a signature
pattern.

Consensus pattern: E-[LIVM]-G-D-K-T-F-[LIVMF](2)-A-
[1] Bairoch A. Unpublished observations (1996).

695. Uncharacterized protein family UPF0021 signature

- 20 The following uncharacterized proteins have been shown [1] to share regions of similarities: -
Yeast chromosome VII hypothetical protein YGL211w. - *Dictyostelium discoideum* protein
veg136. - *Methanococcus jannaschii* hypothetical proteins MJ1157 and MJ1478. These
25 are proteins of from 300 to 360 residues. They can be picked up in the database by the following
pattern which is located in their N-terminal section.

Consensus pattern: C-K-x(2)-F-x(4)-E-x(22,23)-S-G-G-K-D
[1] Bairoch A. Unpublished observations (1997).

696. Uncharacterized protein family UPF0023 signature

30 The following uncharacterized proteins have been shown [1] to share regions of similarities: -
Mouse protein 22A3. - Yeast chromosome XII hypothetical protein YLR022c. -

Caenorhabditis elegans hypothetical protein W06E11.4. - *Methanococcus jannaschii* hypothetical protein MJ0592. These are hydrophilic proteins of about 30 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: D-x-D-E-[LIV]-L-x(4)-V-F-x(3)-S-K-G-

- 5 [1] Bairoch A. Unpublished observations (1997).

697. Uncharacterized protein family UPF0024 signature. The following uncharacterized proteins have been shown [1] to share regions of similarities: - *Escherichia coli* hypothetical protein ygbO and HI0701, the corresponding *Haemophilus influenzae* protein. - *Helicobacter pylori* hypothetical protein HP0926. - Yeast chromosome XV hypothetical protein YOR243c. - *Caenorhabditis elegans* hypothetical protein B0024.11. - *Methanococcus jannaschii* hypothetical proteins MJ0588 and MJ1364. These are hydrophilic proteins of from 39 to 77 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: G-x-K-D-[KR]-x-A-[LV]-T-x-Q-x-[LIVF]-[SGC]-

- [1] Bairoch A. Unpublished observations (1997).

698. Uncharacterized protein family UPF0025 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - *Escherichia coli* hypothetical protein yfcE. - *Bacillus subtilis* hypothetical protein ysnB. - *Mycoplasma genitalium* and *pneumoniae* hypothetical protein MG207. - *Methanococcus jannaschii* hypothetical proteins MJ0623 and MJ0936. These are hydrophilic proteins of about 20 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: D-V-[LIV]-x(2)-G-H-[ST]-H-x(12)-[LIVMF]-N-P-G

- [1] Bairoch A. Unpublished observations (1997).

699. Uncharacterized protein family UPF0029 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Yeast chromosome III hypothetical protein YCR59c. - Yeast chromosome IV hypothetical

protein YDL177C. - *Escherichia coli* hypothetical protein yigZ and HI0722, the corresponding *Haemophilus influenzae* protein. - *Bacillus subtilis* hypothetical protein yvyE. - *A Thermus aquaticus* hypothetical protein in pol 5' region. These proteins can be picked up in the database by the following pattern.

- 5 Consensus pattern: G-x(2)-[LIVM](2)-x(2)-[LIVM]-x(4)-[LIVM]-x(5)-[LIVM](2)-x-R-[FYW](2)-G-G-x(2)-[LIVM]-G

[1] Koonin E.V., Bork P., Sander C. EMBO J. 13:493-503(1994).

- 10 700. Uncharacterized protein family UPF0030 signature

The following uncharacterized proteins have been shown [1] to be highly similar: - Yeast chromosome VI hypothetical protein YFL060c. - Yeast chromosome XIII hypothetical protein YMR095c. - Yeast chromosome XIV hypothetical protein YNL334c. - *Bacillus subtilis* hypothetical protein yaaE. - *Haemophilus influenzae* hypothetical protein HI1648. - *Methanococcus jannaschii* hypothetical protein MJ1661. These are hydrophilic proteins of about 19 to 25 Kd. They can be picked up in the database by the following pattern.

Consensus pattern: [GA]-L-I-[LIV]-P-G-G-E-S-T-[STA]

[1] Bairoch A. Unpublished observations (1997).

- 20 701. Uncharacterized protein family UPF0032 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - *Escherichia coli* hypothetical protein yigU and HI0188, the corresponding *Haemophilus influenzae* protein. - *Bacillus subtilis* hypothetical protein ycbT. - *Mycobacterium tuberculosis* hypothetical protein MtCY49.33c and U2126A, the corresponding *Mycobacterium leprae* protein. - *Synechocystis* strain PCC 6803 hypothetical protein slI0194. - *Odontella sinensis* and *Porphyra purpurea* chloroplast hypothetical protein ycf43. These proteins have from 245 to 317 amino acids and seem to contain at least six or seven transmembrane regions. A conserved region located in the central section of these proteins has been developed as a signature pattern.

Consensus pattern: Y-x(2)-F-[LIVMA](2)-x-L-x(4)-G-x(2)-F-[EQ]-[LIVMF]-P-[LIVM] -

[1] Bairoch A., Rudd K.E. Unpublished observations (1996).

702. Uncharacterized protein family UPF0034 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yhdG and HI0979, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yjbN and HI0634, the corresponding Haemophilus influenzae protein. - Escherichia coli hypothetical protein yohI and HI0270, the corresponding Haemophilus influenzae protein. - Bacillus subtilis hypothetical protein yacF. - Rhodobacter capsulatus protein nifR3 and related proteins in Azospirillum brasilense and Rhizobium leguminosarum. - Synechocystis strain PCC 6803 hypothetical protein slr0644. - Synechocystis strain PCC 6803 hypothetical protein slr0926. - Caenorhabditis elegans hypothetical protein C45G9.2. - Yeast protein SMM1. - Yeast hypothetical protein YLR401c. - Yeast hypothetical protein YLR405w. - Yeast hypothetical protein YML080w. Although it has been proposed [2] that Rhodobacter capsulatus nifR3 is a transcriptional regulatory protein, it is believed that these proteins constitute a family of enzymes whose active site could include a conserved cysteine which has been used as the central part of a signature pattern.

Consensus pattern: [LIVM]-[DNG]-[LIVM]-N-x-G-C-P-x(3)-[LIVMASQ]-x(5)-G-[SAC]
[1] Bairoch A., Rudd K.E. Unpublished observations (1995).[2] Foster-Hartnett D., Cullen P.J., Gabbert K.K., Kranz R.G. Mol. Microbiol. 8:903-914(1993).

703. Uncharacterized protein family UPF0038 signature

The following uncharacterized proteins have been shown [1] to share regions of similarities: - Escherichia coli hypothetical protein yacE and HI0890, the corresponding Haemophilus influenzae protein. - Mycobacterium tuberculosis hypothetical protein MtCY01B2.23 and O410, the corresponding Mycobacterium leprae protein. - Synechocystis strain PCC 6803 hypothetical protein slr0553. - Other hypothetical proteins from Aeromonas hydrophila, Bacteroides nodosus, Neisseria gonorrhoeae, Pseudomonas putida, Thermus thermophilus and Xanthomonas campestris. - Human hypothetical protein pOV-2. - Yeast hypothetical protein YDR196C. - Caenorhabditis elegans hypothetical protein T05G5.5. These proteins all contain, in their N-terminal extremity, an ATP/GTP-binding motif 'A' (P-loop) (see <PDOC00017>). The size of these proteins range from 200 to 290 residues (with the exception of the Mycobacterial sequences which are 410 residues long). A conserved

region some 50 residues away from the ATP-binding P-loop has been developed as a signature pattern.

Consensus pattern: G-x-[LI]-x-R-x(2)-L-x(4)-F-x(8)-[LIV]-x(5)-P-x-[LIV] –

[1] Rudd K.E., Bairoch A. Unpublished observations (1997).

5

704. Ubiquitin-conjugating enzymes active site

Ubiquitin-conjugating enzymes (UBC or E2 enzymes) [1,2,3] catalyze the covalent attachment of ubiquitin to target proteins. An activatedubiquitin moiety is transferred from an ubiquitin-activating enzyme (E1) to E2which later ligates ubiquitin directly to substrate proteins with or without the assistance of 'N-end' recognizing proteins (E3). In most species there are many forms of UBC (at least 9 in yeast) which are implicated in diverse cellular functions. A cysteine residue is required for ubiquitin-thiolester formation. There is a single conserved cysteine in UBC's and the region around that residue is conserved in the sequence of known UBC isozymes. That region has been used as a signature pattern.

10

15

Consensus pattern: [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x- [LIV] [C is the active site residue]

[1] Jentsch S., Seufert W., Sommer T., Reins H.-A. Trends Biochem. Sci. 15:195-198(1990).[2] Jentsch S., Seufert W., Hauser H.-P. Biochim. Biophys. Acta 1089:127-139(1991).[3] Hershko A. Trends Biochem. Sci. 16:265-268(1991).

20

705. Uroporphyrinogen decarboxylase signatures

Uroporphyrinogen decarboxylase (URO-D), the fifth enzyme of the heme biosynthetic pathway, catalyzes the sequential decarboxylation of the four acetyl side chains of uroporphyrinogen to yield coproporphyrinogen [1].URO-D deficiency is responsible for the Human genetic diseases familialporphyria cutanea tarda (fPCT) and hepatoerythropoietic porphyria (HEP).The sequence of URO-D has been well conserved throughout evolution. The best conserved region is located in the N-terminal section; it contains a perfectlyconserved hexapeptide. There are two arginine residues in this hexapeptide which could be involved in the binding, via salt bridges, to the carboxylgroups of the propionate side chains of the substrate. This region has been used as a signature pattern. A second

25

30

signature pattern is based on a another well conserved region which is located in the central section of the protein.

Consensus pattern: P-x-W-x-M-R-Q-A-G-R

Consensus pattern: G-F-[STAGCV]-[STAGC]-x-P-[FYW]-T-[LV]-x(2)-Y-x(2)-[AE]- [GK]

- 5 [1] Garey J.R., Labbe-Bois R., Chelstowska A., Rytka J., Harrison L., Kushner J., Labbe P. Eur. J. Biochem. 205:1011-1016(1992).

706. ubiE/COQ5 methyltransferase family signatures

- 10 The following methyltransferases have been shown [1] to share regions of similarities: - Escherichia coli ubiE, which is involved in both ubiquinone and menaquinone biosynthesis and which catalyzes the S-adenosylmethionine dependent methylation of 2-polyprenyl-6-methoxy-1,4-benzoquinol into 2-polyprenyl-3- methyl-6-methoxy-1,4-benzoquinol and of demethylmenaquinol into menaquinol. - Yeast COQ5, a ubiquinone biosynthesis
15 methyltransferase. - Bacillus subtilis spore germination protein C2 (gene: gercB or gerC2), a probable menaquinone biosynthesis methyltransferase. - Lactococcus lactis gerC2 homolog. - Caenorhabditis elegans hypothetical protein ZK652.9. - Leishmania donovani amastigote-specific protein A41. These are hydrophilic proteins of about 30 Kd (except for ZK652.9 which is 65Kd). They can be picked up in the database by the following patterns.
20 Consensus pattern: Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W
Consensus pattern: R-V-[LIVM]-K-[PV]-G-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S
[1] Lee P.T., Hsu A.Y., Ha H.T., Clarke C.F. J. Bacteriol. 179:1748-1754(1997).

25 707. Uricase signature

Uricase (urate oxidase) [1] is the peroxisomal enzyme responsible for the degradation of urate into allantoin. Some species, like primates and birds, have lost the gene for uricase and are therefore unable to degrade urate. Uricase is a protein of 300 to 400 amino acids. A highly conserved region located in the central part of the sequence has been used as a signature
30 pattern.

Consensus pattern: [LV]-x-[LV]-[LIV]-K-[STV]-[ST]-x-[SN]-x-F-x(2)-[FY]-x(4)- [FY]-x(2)-L-x(5)-R

- [1] Motojima K., Kanaya S., Goto S. J. Biol. Chem. 263:16677-16681(1988).

708. Universal stress protein family (Usp)

By a wide range of stress conditions members of the Usp family are predicted to be related to the MADS-box proteins transcript_fact and bind to DNA [2]. Number of members: 39

[1] Expression and role of the universal stress protein, UspA, of Escherichia coli during growth arrest. Nystrom T, Neidhardt FC; Mol Microbiol 1994; 11:537-544.

[2] Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. Mushegian AR, Koonin EV; Genetics 1996; 144:817-828.

709. Ubiquitin domain signature and profile

Ubiquitin [1,2,3] is a protein of seventy six amino acid residues, found in all eukaryotic cells and whose sequence is extremely well conserved from protozoan to vertebrates. It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis. In most species, there are many genes coding for ubiquitin. However they can be classified into two classes. The first class produces polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin. The number of repeats is variable (up to twelve in a *Xenopus* gene). In the majority of polyubiquitin precursors, there is a final amino-acid after the last repeat. The second class of genes produces precursor proteins consisting of a single copy of ubiquitin fused to a C-terminal extension protein (CEP). There are two types of CEP proteins and both seem to be ribosomal proteins. Ubiquitin is a globular protein, the last four C-terminal residues (Leu-Arg- Gly-Gly) extending from the compact structure to form a 'tail', important for its function. The latter is mediated by the covalent conjugation of ubiquitin to target proteins, by an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine residues in the target proteins. There are a number of proteins which are evolutionary related to ubiquitin: - Ubiquitin-like proteins from baculoviruses as well as in some strains of bovine viral diarrhea viruses (BVDV). These proteins are highly similar to their eukaryotic counterparts. - Mammalian protein GDX [4]. GDX is composed of two domains, a N-terminal ubiquitin-like

710. VHS domain

Domain present in VPS-27, Hrs and STAM. Number of members: 27

711. Vinculin family signatures

Vinculin [1] is a eukaryotic protein that seems to be involved in the attachment of the actin-based microfilaments to the plasma membrane. Vinculin is located at the cytoplasmic side of focal contacts or adhesion plaques. In addition to actin, vinculin interacts with other structural proteins such as talin and alpha-actinins. Vinculin is a large protein of 116 Kd (about a 1000 residues). Structurally the protein consists of an acidic N-terminal domain of about 90 Kd separated from a basic C-terminal domain of about 25 Kd by a proline-rich region of about 50 residues. The central part of the N-terminal domain consists of a variable number (3 in vertebrates, 2 in *Caenorhabditis elegans*) of repeats of a 110 amino acids domain. Catenins [2] are proteins that associate with the cytoplasmic domain of a variety of cadherins. The association of catenins to cadherins produces a complex which is linked to the actin filament network, and which seems to be of primary importance for cadherins cell-adhesion properties. Three different types of catenins seem to exist: alpha, beta, and gamma. Alpha-catenins are proteins of about 100 Kd which are evolutionarily related to vinculin. In terms of their structure the most significant differences are the absence, in alpha-catenin, of the repeated domain and of the proline-rich segment. Two signature patterns for this family of proteins have been developed. The first pattern is located in the N-terminal section of both vinculin and alpha-catenins and is part, in vinculin, of a domain that seems to be involved with the interaction with talin. The second pattern is based on a conserved region in the N-terminal part of the repeated domain of vinculin.

Consensus pattern: [KR]-x-[LIVMF]-x(3)-[LIVMA]-x(2)-[LIVM]-x(6)-R-Q-Q-E-L

Consensus pattern: [LIVM]-x-[QA]-A-x(2)-W-[IL]-x-[DN]-P

[1] Otto J.J. Cell Motil. Cytoskeleton 16:1-6(1990). [2] Herrenknecht K., Ozawa M., Eckerskorn C., Lottspeich F., Lenter M., Kemler R. Proc. Natl. Acad. Sci. U.S.A. 88:9156-9160(1991).

712. (Vitellogenin N) Lipoprotein amino terminal region

This family contains regions from: Vitellogenin, Microsomal triglyceride transfer protein and apolipoprotein B-100. These proteins are all involved in lipid transport [1]. This family contains the LVI chain from lipovitellin, that contains two structural domains.

5 Number of members: 33

[1] The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein.
Anderson TA, Levitt DG, Banaszak LJ Structure 1998;6:895-909.

10 713. (VMSA) Major surface antigen from hepadnavirus

714. ssDNA binding protein (Viral DNA bp)

15 This protein is found in herpesviruses and is needed for replication.

715. (Votage CLC) Voltage gated chloride channels

20 This family of ion channels contains 10 or 12 transmembrane helices. Each protein forms a single pore. It has been shown that some members of this family form homodimers. These proteins contain two CBS domains.

[1] Schmidt-Rose T, Jentsch TJ; J Biol Chem 1997;272:20515-20521.

25 [2] Zhang J, George AL Jr, Griggs RC, Fouad GT, Roberts J, Kwiecinski H, Connolly AM, Ptacek LJ; Neurology 1996;47:993-998.

716. von Willebrand factor type A domain (vwa)

30 More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor.

Bork P, Rohde K;

Biochem J 1991;279:908-911.

1. RUGGERI, Z.M. and WARE, J.

von Willebrand factor.

FASEB J. 7 308-316 (1993).

2. COLOMBATTI, A., BONALDO, P. and DOLIANA, R.

Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins.

MATRIX 13 297-306 (1993).

3. PERKINS, S.J., SMITH, K.F., WILLIAMS, S.C., HARIS, P.I., CHAPMAN, D. and SIM, R.B.

The secondary structure of the von Willebrand factor type A domain in factor B of human complement by Fourier transform infrared spectroscopy. Its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged structure predictions.

J.MOL.BIOL. 238 104-119 (1994).

4. BORK, P. and ROHDE, K.

More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor.

BIOCHEM.J. 279 908-910 (1991).

5. EDWARDS, Y.J.K. and PERKINS, S.J.

The protein fold of the von Willebrand factor type A domain is predicted to be similar to the open twisted beta-sheet flanked by alpha-helices found in human ras-p21.

FEBS LETT. 358 283-286 (1995).

6. LEE, J.O., RIEU, P., ARNAOUT, M.A. and LIDDINGTON, R.

Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18).
CELL 80 631-638 (1995).

5 7. QU, A. and LEAHY, D.J.

Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha L beta 2) integrin.
PROC.NATL.ACAD.SCI.USA 92 10277-10281 (1995).

10 The von Willebrand factor is a large multimeric glycoprotein found in blood plasma. Mutant forms are involved in the aetiology of bleeding disorders [1]. In von Willebrand factor, the type A domain (vWF) is the prototype for a protein superfamily. The vWF domain is found in various plasma proteins: complement factors B, C2, CR3 and CR4; the integrins (I-domains); collagen types VI, VII, XII and XIV; and other extracellular proteins [2-4]. Proteins that incorporate vWF domains participate in numerous biological events (e.g., cell adhesion, migration, homing, pattern formation, and signal transduction), involving interaction with a large array of ligands [2].
15 Secondary structure prediction from 75 aligned vWF sequences has revealed a largely alternating sequence of alpha-helices and beta-strands [3]. Fold recognition algorithms were used to score sequence compatibility with a library of known structures: the vWF domain fold was predicted to be a doubly-wound, open, twisted beta-sheet flanked by alpha-helices [5].
20 3D structures have been determined for the I-domains of integrins CD11b (with bound magnesium) [6] and CD11a (with bound manganese) [7]. The domain adopts a classic alpha/beta Rossmann fold and contains an unusual metal ion coordination site at its surface. It has been suggested that this site represents a general metal ion-dependent adhesion site (MIDAS) for binding protein ligands [6]. The residues constituting the MIDAS motif in the CD11b
25 and CD11a I-domains are completely conserved, but the manner in which the metal ion is coordinated differs slightly [7].
30

VWFADOMAIN is a 3-element fingerprint that provides a signature for the vWF

domain superfamily. The fingerprint was derived from an initial alignment of 14 sequences. Motif 1 includes the first beta-strand and 3 conserved residues involved in metal ion coordination in I-domains (Asp and 2 serines in positions 8, 10 and 12, respectively); motif 2 spans strands beta-2 and beta-2'; and motif 3 encodes beta-strand 3 and a conserved Asp (in position 7), which coordinates the metal ion [6,7]. Three iterations on OWL27.0 were required to reach convergence, at which point a true set comprising 56 sequences was identified. Numerous partial matches were also found.

717. (WD40) WD domain, G-beta repeat

The ancient regulatory-protein family of WD-repeat proteins.

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF;

Nature 1994;371:297-300.

Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors [1]. The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

In higher eukaryotes G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). Such a repetitive segment has been shown [E1,2,3,4,5] to exist in a number of other proteins listed below:

- Yeast STE4, a component of the pheromone response pathway. STE4 is a G-beta like protein that associates with GPA1 (G-alpha) and STE18 (G-gamma).
- Yeast MS11, a negative regulator of RAS-mediated cAMP synthesis. MS11 is most probably also a G-beta protein.

- Human and chicken protein 12.3. The function of this protein is not known, but on the basis of its similarity to G-beta proteins, it may also function in signal transduction.
- 5 - *Chlamydomonas reinhardtii* gblp. This protein is most probably the homolog of vertebrate protein 12.3.
- Human LIS1, a neuronal protein involved in type-1 lissencephaly [E2].
- Mammalian coatomer beta' subunit (beta'-COP), a component of a cytosolic protein complex that reversibly associates with Golgi membranes to form vesicles that mediate biosynthetic protein transport.
- 10 - Yeast CDC4, essential for initiation of DNA replication and separation of the spindle pole bodies to form the poles of the mitotic spindle.
- Yeast CDC20, a protein required for two microtubule-dependent processes: nuclear movements prior to anaphase and chromosome separation.
- 15 - Yeast MAK11, essential for cell growth and for the replication of M1 double-stranded RNA.
- Yeast PRP4, a component of the U4/U6 small nuclear ribonucleoprotein with a probable role in mRNA splicing.
- 20 - Yeast PWP1, a protein of unknown function.
- Yeast SKI8, a protein essential for controlling the propagation of double-stranded RNA.
- Yeast SOF1, a protein required for ribosomal RNA processing which associates with U3 small nucleolar RNA.
- 25 - Yeast TUP1 (also known as AER2 or SFL2 or CYC9), a protein which has been implicated in dTMP uptake, catabolite repression, mating sterility, and many other phenotypes.
- Yeast YCR57c, an ORF of unknown function from chromosome III.
- Yeast YCR72c, an ORF of unknown function from chromosome III.
- 30 - Slime mold coronin, an actin-binding protein.
- Slime mold AAC3, a developmentally regulated protein of unknown function.

- Drosophila protein Groucho (formerly known as E(spl); 'enhancer of split'), a protein involved in neurogenesis and that seems to interact with the Notch and Delta proteins.
- Drosophila TAF-II-80, a protein that is tightly associated with TFIID.

The number of repeats in the above proteins varies between 5 (PRP4, TUP1, and Groucho) and 8 (G-beta, STE4, MS11, AAC3, CDC4, PWP1, etc.). In G-beta and G-beta like proteins, the repeats span the entire length of the sequence, while in other proteins, they make up the N-terminal, the central or the C-terminal section.

A signature pattern can be developed from the central core of the domain (positions 9 to 23).

-Consensus pattern: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN]

[1] Gilman A.G.

Annu. Rev. Biochem. 56:615-649(1987).

[2] Duronio R.J., Gordon J.I., Boguski M.S.

Proteins 13:41-56(1992).

[3] van der Voorn L., Ploegh H.L.

FEBS Lett. 307:131-134(1992).

[4] Neer E.J., Schmidt C.J., Nambudripad R., Smith T.F.

Nature 371:297-300(1994).

[5] Smith T.F., Gaiatzes C.G., Saxena K., Neer E.J.

Biochemistry In Press(1998).

718. WHEP-TRS domain containing proteins

A conserved domain of 46 amino acids has been shown [1] to exist in a number of higher eukaryote aminoacyl-transfer RNA synthetases. This domain is present

one to six times in the following enzymes:

- Mammalian multifunctional aminoacyl-tRNA synthetase. The domain is present three times in a region that separates the N-terminal glutamyl-tRNA synthetase domain from the C-terminal prolyl-tRNA synthetase domain.
- Drosophila multifunctional aminoacyl-tRNA synthetase. The domain is present six times in the intercatalytic region.
- Mammalian tryptophanyl-tRNA synthetase. The domain is found at the N-terminal extremity.
- Mammalian, insect, nematode and plant glycyl-tRNA synthetase. The domain is found at the N-terminal extremity [2].
- Mammalian histidyl-tRNA synthetase. The domain is found at the N-terminal extremity.

This domain, which is called WHEP-TRS, could contain a central alpha-helical region and may play a role in the association of tRNA-synthetases into multienzyme complexes.

A signature pattern based on the first 29 positions of the WHEP-Domain has been developed.

-Consensus pattern: [QY]-G-[DNEA]-x-[LIV]-[KR]-x(2)-K-x(2)-[KRNG]-[AS]-x(4)-[LIV]-[DENK]-x(2)-[IV]-x(2)-L-x(3)-K

- [1] Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.
EMBO J. 10:4267-4277(1991).
- [2] Nada S., Chang P.K., Dignam J.D.
J. Biol. Chem. 268:7660-7667(1993).

719. (Worm family 8) Putative membrane protein
Analysis of protein domain families in *Caenorhabditis elegans*.
Sonnhammer EL, Durbin R;

Genomics 1997;46:200-216.

This family called family 8 in [1], may be a transmembrane protein

The specific function of this protein is unknown.

5

720. Xylose isomerase

Xylose isomerase (EC 5.3.1.5) [1] is an enzyme found in microorganisms which catalyzes the interconversion of D-xylose to D-xylulose. It can also isomerize D-ribose to D-ribulose and D-glucose to D-fructose. Xylose isomerase seems to require magnesium for its activity, while cobalt is necessary to stabilize the tetrameric structure of the enzyme. A number of residues are conserved in all known xylose isomerases.

10

Xylose isomerase also exists in plants [2] where it is homodimeric and is manganese-dependent.

15

Two signature patterns for xylose isomerase have been developed. The first one is derived from a stretch of five conserved amino acids that includes a glutamic acid residue known to be one of the four residues involved in the binding of the magnesium ion [3]; this pattern also includes a lysine residue which is involved in the catalytic activity. The second pattern is derived from a conserved region in the N-terminal section of the enzyme that includes an histidine residue which has been shown [4] to be involved in the catalytic mechanism of the enzyme.

20

25

-Consensus pattern: [LI]-E-P-K-P-x(2)-P

[E is a magnesium ligand]

[K is an active site residue]

-Consensus pattern: [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE]

30

[H is an active site residue]

[1] Dauter Z., Dauter M., Hemker J., Witzel H., Wilson K.S.

FEBS Lett. 247:1-8(1989).

[2] Kristo P.A., Saarelainen R., Fagerstrom R., Aho S., Korhola M.

Eur. J. Biochem. 237:240-246(1996).

[3] Henrick K., Collyer C.A., Blow D.M.

J. Mol. Biol. 208:129-157(1989).

5 [4] Vangrysperre W., Ampe C., Kersters-Hilderson H., Tempst P.

Biochem. J. 263:195-199(1989).

721. XPG protein signatures. Xeroderma pigmentosum (XP) [1] is a human autosomal
 10 recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's
 skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the
 incision step of DNA excision repair. There are a minimum of seven genetic
 complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can
 be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2]. XPG belongs to a family
 15 of proteins [2,3,4,5,6] that are composed of two main subsets: - Subset 1, to which belongs
 XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-
 stranded DNA endonucleases [7,8]. XPG makes the 3' incision in human DNA nucleotide
 excision repair [9]. - Subset 2, to which belongs mouse and human FEN-1, rad2 from fission
 yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease. In
 20 addition to the proteins listed in the above groups, this family also includes: - Fission yeast
 exo1, a 5'→3' double-stranded DNA exonuclease that could act in a pathway that corrects
 mismatched base pairs. - Yeast EXO1 (DHS1), a protein with probably the same function as
 exo1. - Yeast DIN7. Sequence alignment of this family of proteins reveals that similarities are
 largely confined to two regions. The first is located at the N-terminal extremity (N-region)
 25 and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region)
 and found towards the C-terminus; it spans about 140 residues and contains a highly
 conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]).
 It is possible that the conserved acidic residues are involved in the catalytic mechanism of
 DNA excision repair in XPG. The amino acids linking the N- and I-regions are not
 30 conserved; indeed, they are largely absent from proteins belonging to the second subset. Two
 signature patterns have been developed for these proteins. The first corresponds to the central
 part of the N-region, the second to part of the I-region and includes the putative catalytic core
 pentapeptide

Consensus pattern: [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K-

Consensus pattern: [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-
[CLM]-

- [1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).[2] Scherly D., Nouspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993).[3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993).[4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994).[5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994).[6] Szankasi P., Smith G.R. Science 267:1166-1169(1995).[7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993).[8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).

722. Xanthine/uracil permeases family

The following transport proteins which are involved in the uptake of xanthine or uracil are evolutionary related [1]:

- Uric acid-xanthine permease (gene uapA) from *Aspergillus nidulans*.
- Purine permease (gene uapC) from *Aspergillus nidulans*.
- Xanthine permease from *Bacillus subtilis* (gene pbuX).
- Uracil permease from *Escherichia coli* (gene uraA) [2] and *Bacillus* (gene pyrP).
- Hypothetical protein yedG from *Escherichia coli*.
- Hypothetical protein ygfO from *Escherichia coli*.
- Hypothetical protein ygfU from *Escherichia coli*.
- Hypothetical protein yicE from *Escherichia coli*.
- Hypothetical protein yunJ from *Bacillus subtilis*.
- Hypothetical protein yunK from *Bacillus subtilis*.

They are proteins of from 430 to 595 residues that seem to contain 12 transmembrane domains.

The best conserved region which corresponds with what seems to be the tenth transmembrane domain has been selected as a signature pattern.

5

-Consensus pattern: [LIVM]-P-x-[PASIF]-V-[LIVM]-G-G-x(4)-[LIVM]-[FY]-[GSA]-x-[LIVM]-x(3)-G

[1] Diallynas G., Gorfinkiel L., Arst G., Cecchetto G., Scazzocchio C.

J. Biol. Chem. 270:8610-8622(1995).

10

[2] Andersen P.S., Frees D., Fast R., Mygind B.

J. Bacteriol. 177:2008-2013(1995).

723. Hypothetical yabO/yceC/sfhB family

15

The following proteins, which seems to belong to a family of pseudouridine synthases (EC 4.2.1.70) [1] have been shown to share regions of similarities:

20

- *Escherichia coli* and *Haemophilus influenzae* ribosomal large subunit pseudouridine synthase A (gene *rluA*). It is responsible for synthesis of pseudouridine from uracil-746 IN 23S rRNA.

- *Escherichia coli* and *Haemophilus influenzae* ribosomal large subunit pseudouridine synthase C (gene *rluC*). It is responsible for synthesis of pseudouridine from uracil at positions 955, 2504 and 2580 in 23S rRNA.

25

- *Escherichia coli* protein and homologs in other bacteria large subunit pseudouridine synthase D (gene *rluD*).

- Yeast DRAP deaminase (gene *RIB2*).

- *Escherichia coli* hypothetical protein *yqcB* and HI1435, the corresponding *Haemophilus influenzae* protein.

- *Haemophilus influenzae* hypothetical protein HI0042.

30

- *Aquifex aeolicus* hypothetical protein AQ_1758.

- *Bacillus subtilis* hypothetical protein *yhcT*.

- *Bacillus subtilis* hypothetical protein *yjbO*.

- *Bacillus subtilis* hypothetical protein *glyB*.

574

- *Helicobacter pylori* hypothetical protein HP0347.
- *Helicobacter pylori* hypothetical protein HP0745.
- *Helicobacter pylori* hypothetical protein HP0956.
- *Mycoplasma genitalium* hypothetical protein MG209.
- 5 - *Mycoplasma genitalium* hypothetical protein MG370.
- *Synechocystis* strain PCC 6803 hypothetical protein slr1592.
- *Synechocystis* strain PCC 6803 hypothetical protein slr1629.
- Yeast hypothetical protein YDL036c.
- Yeast hypothetical protein YGR169c.
- 10 - Fission yeast hypothetical protein SpAC18B11.02c.
- *Caenorhabditis elegans* hypothetical protein K07E8.7.

These are proteins of from 21 to 50 Kd which contain a number of conserved regions in their central section. They can be picked up in the database by the following highly conserved pattern.

-Consensus pattern: [LIVCA]-[NHYT]-R-[LI]-D-x(2)-T-[STA]-G-[LIVAGC]-
[LIVMF](2)-[LIVMFGC]-[SGTACV]

[1] Conrad J., Sun D., Englund N., Ofengand J.
J. Biol. Chem. 273:18562-18566(1998).

In addition, the following bacterial proteins, which seems to belong to a family of pseudouridine synthases (EC 4.2.1.70) [1] also have been shown to share regions of similarities:

- *Escherichia coli* and *Haemophilus influenzae* 16S pseudouridylylate 516 synthase (EC 4.2.1.70) (gene: *rsuA*). This enzyme is responsible for the formation of pseudouridine from uracil-516 in 16S ribosomal RNA.
- 30 - *Escherichia coli* hypothetical protein *yciL* and HI1199, the corresponding *Haemophilus influenzae* protein.
- *Escherichia coli* hypothetical protein *yjbC*.
- *Escherichia coli* hypothetical protein *ymfC* and HI0694, the corresponding

575

Haemophilus influenzae protein.

- Aquifex aeolicus hypothetical protein AQ_554.
- Aquifex aeolicus hypothetical protein AQ_1464.
- Bacillus subtilis hypothetical protein ypuL.
- 5 - Bacillus subtilis hypothetical protein ytzF.
- Borrelia burgdorferi hypothetical protein BB0129.
- Helicobacter pylori hypothetical protein HP1459.
- Synechocystis strain PCC 6803 hypothetical protein slr0361.
- Synechocystis strain PCC 6803 hypothetical protein slr0612.

10

These are proteins of from 25 to 40 Kd which contain a number of conserved regions in their central section. They can be picked up in the database by the following highly conserved pattern.

15

-Consensus pattern: G-R-L-D-x(2)-[STA]-x-G-[LIVFA]-[LIVMF](3)-[ST]-[DNST]

[1] Wrzesinski J., Bakin A., Nurse K., Lane B.G., Ofengand J.
 Biochemistry 34:8904-8913(1995).

20

724. Zinc finger present in dystrophin, CBP/p300
 ZZ in dystrophin binds calmodulin
 Putative zinc finger; binding not yet shown.

25

725. Zinc carboxypeptidase

There are a number of different types of zinc-dependent carboxypeptidases (EC 3.4.17.-) [1,2]. All these enzymes seem to be structurally and functionally related. The enzymes that belong to this family are listed below.

30

- Carboxypeptidase A1 (EC 3.4.17.1), a pancreatic digestive enzyme that can removes all C-terminal amino acids with the exception of Arg, Lys and Pro.
- Carboxypeptidase A2 (EC 3.4.17.15), a pancreatic digestive enzyme with a

specificity similar to that of carboxypeptidase A1, but with a preference for bulkier C-terminal residues.

- Carboxypeptidase B (EC 3.4.17.2), also a pancreatic digestive enzyme, but that preferentially removes C-terminal Arg and Lys.
- 5 - Carboxypeptidase N (EC 3.4.17.3) (also known as arginine carboxypeptidase), a plasma enzyme which protects the body from potent vasoactive and inflammatory peptides containing C-terminal Arg or Lys (such as kinins or anaphylatoxins) which are released into the circulation.
- Carboxypeptidase H (EC 3.4.17.10) (also known as enkephalin convertase or 10 carboxypeptidase E), an enzyme located in secretory granules of pancreatic islets, adrenal gland, pituitary and brain. This enzyme removes residual C-terminal Arg or Lys remaining after initial endoprotease cleavage during prohormone processing.
- Carboxypeptidase M (EC 3.4.17.12), a membrane bound Arg and Lys specific 15 enzyme.
It is ideally situated to act on peptide hormones at local tissue sites where it could control their activity before or after interaction with specific plasma membrane receptors.
- Mast cell carboxypeptidase (EC 3.4.17.1), an enzyme with a specificity 20 to carboxypeptidase A, but found in the secretory granules of mast cells.
- *Streptomyces griseus* carboxypeptidase (Cpase SG) (EC 3.4.17.-) [3], which combines the specificities of mammalian carboxypeptidases A and B.
- *Thermoactinomyces vulgaris* carboxypeptidase T (EC 3.4.17.18) (CPT) [4], which also combines the specificities of carboxypeptidases A and B.
- 25 - AEBP1 [5], a transcriptional repressor active in preadipocytes. AEBP1 seems to regulate transcription by cleavage of other transcriptional proteins.
- Yeast hypothetical protein YHR132c.

All of these enzymes bind an atom of zinc. Three conserved residues are 30 implicated in the binding of the zinc atom: two histidines and a glutamic acid
Two signature patterns which contain these three zinc-ligands have been derived.

-Consensus pattern: [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-

[STAG]-x(6)-[LIVMFYTA]

[H and E are zinc ligands]

-Consensus pattern: H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW]

[H is a zinc ligand]

5

[1] Tan F., Chan S.J., Steiner D.F., Schilling J.W., Skidgel R.A.

J. Biol. Chem. 264:13165-13170(1989).

[2] Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,

Serafin W.E.

10 J. Biol. Chem. 264:20094-20099(1989).

[3] Narahashi Y.

J. Biochem. 107:879-886(1990).

[4] Teplyakov A., Polyakov K., Obmolova G., Strokopytov B., Kuranova I.,

Osterman A.L., Grishin N.V., Smulevitch S.V., Zagnitko O.P.,

15 Galperina O.V., Matz M.V., Stepanov V.M.

Eur. J. Biochem. 208:281-288(1992).

[5] He G.-P., Muise A., Li A.W., Ro H.-S.

Nature 378:92-96(1995).

[6] Hourdou M.-L., Guinand M., Vacheron M.J., Michel G., Denoroy L.,

20 Duez C.M., Englebert S., Joris B., Weber G., Ghuysen J.-M.

Biochem. J. 292:563-570(1993).

[7] Rawlings N.D., Barrett A.J.

Meth. Enzymol. 248:183-228(1995).

25

726. Zinc finger, C2H2 type

The C2H2 zinc finger is the classical zinc finger domain.

The two conserved cysteines and histidines co-ordinate a

zinc ion. The following pattern describes the zinc finger.

30 #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C]

Where X can be any amino acid, and numbers in brackets

indicate the number of residues. The positions marked # are

those that are important for the stable fold of the zinc

finger. The final position can be either his or cys.

The C2H2 zinc finger is composed of two short beta strands followed by an alpha helix. The amino terminal part of the helix binds the major groove in DNA binding zinc fingers.

5

'Zinc finger' domains [1-5] are nucleic acid-binding protein structures first identified in the *Xenopus* transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:

10

15

x x

x x

x x

x x

x x

20

x x

C H

x \ / x

x Zn x

x / \ x

25

C H

x x x x x x x x x x

Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

30

Some of the proteins known to include C2H2-type zinc fingers are listed below. The number of zinc finger regions found in each of these proteins are indicated between brackets; a '+' symbol indicates that only partial sequence

5 data is available and that additional finger domains may be present.

- *Saccharomyces cerevisiae*: ACE2 (3), ADR1 (2), AZF1 (4), FZF1 (5), MIG1 (2), MSN2 (2), MSN4 (2), RGM1 (2), RIM1 (3), RME1 (3), SFP1 (2), SSL1 (1), STP1 (3), SWI5 (3), VAC1 (1) and ZMS1 (2).

10 - *Emmericella nidulans*: brlA (2), creA (2).

- *Drosophila*: AEF-1 (4), Cf2 (7), ci-D (5), Disconnected (2), Escargot (5), Glass (5), Hunchback (6), Kruppel (5), Kruppel-H (4+), Odd-skipped (4), Odd-paired (4), Pep (3), Snail (5), Spalt-major (7), Serendipity locus beta (6), delta (7), h-1 (8), Suppressor of hairy wing su(Hw) (12), Suppressor of variegation suvar(3)7 (5), Teashirt (3) and Tramtrack (2).

15 - *Xenopus*: transcription factor TFIIIA (9), p43 from RNP particle (9), Xfin (37 !!), Xsna (5), gastrula XlclGF5.1 to XlclGF71.1 (from 4+ to 11+), Oocyte XlclOF2 to XlclOF22 (from 7 to 12).

20 - Mammalian: basonuclin (6), BCL-6/LAZ-3 (6), erythroid krueppel-like transcription factor (3), transcription factors Sp1 (3), Sp2 (3), Sp3 (3) and Sp(4)3, transcriptional repressor YY1 (4), Wilms' tumor protein (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), Zfx (13), Zfy (13), Zfp-35 (18), ZNF7 (15), ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3).

30 In addition to the conserved zinc ligand residues it has been shown [6] that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue.

-Consensus pattern: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

[The two C's and two H's are zinc ligands]

[1] Klug A., Rhodes D.

Trends Biochem. Sci. 12:464-469(1987).

5 [2] Evans R.M., Hollenberg S.M.

Cell 52:1-3(1988).

[3] Payre F., Vincent A.

FEBS Lett. 234:245-250(1988).

[4] Miller J., McLachlan A.D., Klug A.

10 EMBO J. 4:1609-1614(1985).

[5] Berg J.M.

Proc. Natl. Acad. Sci. U.S.A. 85:99-102(1988).

[6] Rosenfeld R., Margalit H.

J. Biomol. Struct. Dyn. 11:557-570(1993).

727. Zinc finger, C3HC4 type (RING finger)

A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or 'RING' finger) [1] that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D structure of the zinc ligation system is unique to the RING domain and is referred to as the "cross-brace" motif. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C.

Proteins currently known to include the C3HC4 domain are listed below (references are only provided for recently determined sequences).

- Mammalian V(DJ) recombination activating protein (gene RAG1). RAG1 activates the rearrangement of immunoglobulin and T-cell receptor genes.
- Mouse rpt-1. Rpt-1 is a trans-acting factor that regulates gene expression directed by the promoter region of the interleukin-2 receptor alpha chain or the LTR promoter region of HIV-1.

- Human rfp. Rfp is a developmentally regulated protein that may function in male germ cell development. Recombination of the N-terminal section of rfp with a protein tyrosine kinase produces the ret transforming protein.
- Human 52 Kd Ro/SS-A protein. A protein of unknown function from the Ro/SS-A ribonucleoprotein complex. Sera from patients with systemic lupus erythematosus or primary Sjogren's syndrome often contain antibodies that react with the Ro proteins.
- Human histocompatibility locus protein RING1.
- Human PML, a probable transcription factor. Chromosomal translocation of PML with retinoic receptor alpha creates a fusion protein which is the cause of acute promyelocytic leukemia (APL).
- Mammalian breast cancer type 1 susceptibility protein (BRCA1) [E1].
- Mammalian cbl proto-oncogene.
- Mammalian bmi-1 proto-oncogene.
- Vertebrate CDK-activating kinase (CAK) assembly factor MAT1, a protein that stabilizes the complex between the CDK7 kinase and cyclin H (MAT1 stands for 'Menage A Trois').
- Mammalian mel-18 protein. Mel-18 which is expressed in a variety of tumor cells is a transcriptional repressor that recognizes and bind a specific DNA sequence.
- Mammalian peroxisome assembly factor-1 (PAF-1) (PMP35), which is somewhat involved in the biogenesis of peroxisomes. In humans, defects in PAF-1 are responsible for a form of Zellweger syndrome, an autosomal recessive disorder associated with peroxisomal deficiencies.
- Human MAT1 protein, which interacts with the CDK7-cyclin H complex.
- Human RING1 protein.
- Xenopus XNF7 protein, a probable transcription factor.
- Trypanosoma protein ESAG-8 (T-LR), which may be involved in the posttranscriptional regulation of genes in VSG expression sites or may interact with adenylate cyclase to regulate its activity.
- Drosophila proteins Posterior Sex Combs (Psc) and Suppressor two of zeste (Su(2)2). The two proteins belong to the Polycomb group of genes needed to maintain the segment-specific repression of homeotic selector genes.

582

- Drosophila protein male-specific msl-2, a DNA-binding protein which is involved in X chromosome dosage compensation (the elevation of transcription of the male single X chromosome).
- Arabidopsis thaliana protein COP1 which is involved in the regulation of photomorphogenesis.
- Fungal DNA repair proteins RAD5, RAD16, RAD18 and rad8.
- Herpesviruses trans-acting transcriptional protein ICP0/IE110. This protein which has been characterized in many different herpesviruses is a trans-activator and/or -repressor of the expression of many viral and cellular promoters.
- Baculoviruses protein CG30.
- Baculoviruses major immediate early protein (PE-38).
- Baculoviruses immediate-early regulatory protein IE-N/IE-2.
- Caenorhabditis elegans hypothetical proteins F54G8.4, R05D3.4 and T02C1.1.
- Yeast hypothetical proteins YER116c and YKR017c.

The central region of the domain was selected as a signature pattern for the C3HC4 finger.

-Consensus pattern: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA]

[1] Borden K.L.B., Freemont P.S.

Curr. Opin. Struct. Biol. 6:395-401(1996).

728. Zinc finger C-x8-C-x5-C-x3-H type (and similar).

729. Zinc finger, CCHC class

A family of CCHC zinc fingers, mostly from retroviral gag proteins (nucleocapsid). Prototype structure is from HIV.

Also contains members involved in eukaryotic gene regulation, such as C. elegans GLH-1.

Structure is an 18-residue zinc finger; no examples of indels in the alignment.

5 730. Zn-finger in Ran binding protein and others.

731. AN1-like Zinc finger

10 Zinc finger at the C-terminus of An1 [Swiss:Q91889](#), a ubiquitin-like protein in *Xenopus laevis*. The following pattern describes the zinc finger. C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C Where X can be any amino acid, and numbers in brackets indicate the number of residues.

15 [1] Linnen JM, Bailey CP, Weeks DL; Gene 1993;128:181-188.

732. 14-3-3 proteins

Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways.

Xiao B, Smerdon SJ, Jones DH, Dodson GG, Soneji Y, Aitken A, Gamblin SJ; Nature 1995;376:188-191.

Crystal structure of the zeta isoform of the 14-3-3 protein.

Liu D, Bienkowska J, Petosa C, Collier RJ, Fu H, Liddington R;

Nature 1995;376:191-194.

Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine.

Muslin AJ, Tanner JW, Allen PM, Shaw AS;

Cell 1996;84:889-897.

The 14-3-3 protein binds its target proteins with a common site located towards the C-terminus.

Ichimura T, Ito M, Itagaki C, Takahashi M, Horigome T, Omata S, Ohno S,
Isobe T
FEBS Lett 1997;413:273-276.

- 5 Molecular evolution of the 14-3-3 protein family.

Wang W, Shakes DC

J Mol Evol 1996;43:384-398.

Function of 14-3-3 proteins.

Jin DY, Lyu MS, Kozak CA, Jeang KT

- 10 Nature 1996;382:308-308.

The 14-3-3 proteins [1,2,3] are a family of closely related acidic homodimeric proteins of about 30 Kd which were first identified as being very abundant in mammalian brain tissues and located preferentially in neurons. The 14-3-3 proteins seem to have multiple biological activities and play a key role in signal transduction pathways and the cell cycle. They interact with kinases such as PKC or Raf-1; they seem to also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases and in plants they are associated with a complex that binds to the G-box promoter elements.

The 14-3-3 family of proteins are ubiquitously found in all eukaryotic species studied and have been sequenced in fungi (yeast BMH1 and BMH2, fission yeast rad24 and rad25), plants, Drosophila, and vertebrates. The sequences of the 14-3-3 proteins are extremely well conserved. Two highly conserved regions have been selected as signature patterns: the first is a peptide of 11 residues located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section.

-Consensus pattern: R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA]

-Consensus pattern: Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD]

[1] Aitken A.

Trends Biochem. Sci. 20:95-97(1995).

[2] Morrison D.

Science 266:56-57(1994).

[3] Xiao B., Smerdon S.J., Jones D.H., Dodson G.G., Soneji Y., Aitken A.,

5 Gamblin S.J.

Nature 376:188-191(1995).

733. D-isomer specific 2-hydroxyacid dehydrogenases (2 Hacid DH)

10

This Pfam covers the Formate dehydrogenase, D-glycerate dehydrogenase and D-lactate dehydrogenase families in SCOP. A number of NAD-dependent 2-hydroxyacid dehydrogenases which seem to be specific for the D-isomer of their substrate have been shown [1,2,3,4] to be functionally and structurally related. These enzymes are listed below.

15

- D-lactate dehydrogenase (EC 1.1.1.28), a bacterial enzyme which catalyzes the reduction of D-lactate to pyruvate.

- D-glycerate dehydrogenase (EC 1.1.1.29) (NADH-dependent hydroxypyruvate reductase), a plant leaf peroxisomal enzyme that catalyzes the reduction of hydroxypyruvate to glycerate. This reaction is part of the glycolate pathway of photorespiration.

20

- D-glycerate dehydrogenase from the bacteria *Hyphomicrobium methylovorum* and *Methylobacterium extorquens*.

- 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), a bacterial enzyme that catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. This reaction is the first committed step in the 'phosphorylated' pathway of serine biosynthesis.

25

- Erythronate-4-phosphate dehydrogenase (EC 1.1.1.-) (gene *pdxB*), a bacterial enzyme involved in the biosynthesis of pyridoxine (vitamin B6).

- D-2-hydroxyisocaproate dehydrogenase (EC 1.1.1.-) (D-hicDH), a bacterial enzyme that catalyzes the reversible and stereospecific interconversion between 2-ketocarboxylic acids and D-2-hydroxy-carboxylic acids.

30

- Formate dehydrogenase (EC 1.2.1.2) (FDH) from the bacteria *Pseudomonas* sp. 101 and various fungi [5].

- Vancomycin resistance protein *vanH* from *Enterococcus faecium*; this protein is a D-specific alpha-keto acid dehydrogenase involved in the formation of a

peptidoglycan which does not terminate by D-alanine thus preventing vancomycin binding.

- Escherichia coli hypothetical protein ycdW.
- Escherichia coli hypothetical protein yiaE.
- 5 - Haemophilus influenzae hypothetical protein HI1556.
- Yeast hypothetical protein YER081w.
- Yeast hypothetical protein YIL074w.

All these enzymes have similar enzymatic activities and are structurally related. Three of the most conserved regions of these proteins have been selected to develop patterns. The first pattern is based on a glycine-rich region located in the central section of these enzymes; this region probably corresponds to the NAD-binding domain. The two other patterns contain a number of conserved charged residues, some of which may play a role in the catalytic mechanism.

- 15 -Consensus pattern: [LIVMA]-[AG]-[IVT]-[LIVMFY]-[AG]-x-G-[NHKROGSAC]-[LIV]-G-x(13,14)-[LIVfMT]-x(2)-[FYwCTH]-[DNSTK]
- Consensus pattern: [LIVMFYWA]-[LIVFYWC]-x(2)-[SAC]-[DNQHR]-[IVFA]-[LIVF]-x-[LIVF]-[HNI]-x-P-x(4)-[STN]-x(2)-[LIVMF]-x-[GSDN]
- Consensus pattern: [LMFATC]-[KPQ]-x-[GSTDN]-x-[LIVMFYWR]-[LIVMFYW](2)-N-x-[STAGC]-R-[GP]-x-[LIVH]-[LIVMC]-[DNV]
- 20

[1] Grant G.A. Biochem. Biophys. Res. Commun. 165:1371-1374(1989).

[2] Kochhar S., Hunziker P., Leong-Morgenthaler P.M., Hottinger H. Biochem. Biophys. Res. Commun. 184:60-66(1992).

25 [3] Ohta T., Taguchi H. J. Biol. Chem. 266:12588-12594(1991).

[4] Goldberg J.D., Yoshida T., Brick P. J. Mol. Biol. 236:1123-1140(1994).

[5] Popov V.O., Lamzin V.S. Biochem. J. 301:625-643(1994).

30 734. 2-oxo acid dehydrogenases acyltransferase (catalytic domain)
Refined crystal structure of the catalytic domain of dihydrolipoyl transacetylase (E2P) from azotobacter vinelandii at 2.6 angstroms resolution.

Mattevi A, Obmolova G, Kalk KH, Westphal AH, De Kok A, Hol WG;
J Mol Biol 1993;230:1183-1199.

These proteins contain one to three copies of a lipoyl binding domain
followed by the catalytic domain.

5

735. 3-beta hydroxysteroid dehydrogenase/isomerase family

Structure and tissue-specific expression of 3

beta-hydroxysteroid dehydrogenase/5-ene-4-ene isomerase

10 genes in human and rat classical and peripheral

steroidogenic tissues.

Labrie F, Simard J, Luu-The V, Pelletier G, Belanger A,

Lachance Y, Zhao HF, Labrie C, Breton N, de Launoit Y, et al

J Steroid Biochem Mol Biol 1992;41:421-435.

15 The enzyme 3 beta-hydroxysteroid dehydrogenase/5-ene-4-ene

isomerase (3 beta-HSD) catalyzes the oxidation and isomerization

of 5-ene-3 beta-hydroxypregnene and 5-ene-hydroxyandrostene

steroid precursors into the corresponding 4-ene-ketosteroids necessary

for the formation of all classes of steroid hormones.

20

736. 3-hydroxyacyl-CoA dehydrogenase

This family also includes lambda crystallin.

Structure of L-3-hydroxyacyl-coenzyme A dehydrogenase:

25 preliminary chain tracing at 2.8-A resolution.

Birktoft JJ, Holden HM, Hamlin R, Xuong NH, Banaszak LJ;

Proc Natl Acad Sci U S A 1987;84:8262-8266.

3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) (HCDH) [1] is an enzyme involved

30 in fatty acid metabolism, it catalyzes the reduction of 3-hydroxyacyl-CoA to

3-oxoacyl-CoA. Most eukaryotic cells have 2 fatty-acid beta-oxidation systems,

one located in mitochondria and the other in peroxisomes. In peroxisomes

3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA hydratase (ECH) and

3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme where the N-terminal domain bears the hydratase/isomerase activities and the C-terminal domain the dehydrogenase activity. There are two mitochondrial enzymes: one which is monofunctional and the other which is, like its peroxisomal counterpart, multifunctional.

In *Escherichia coli* (gene *fadB*) and *Pseudomonas fragi* (gene *faoA*) HCDH is part of a multifunctional enzyme which also contains an ECH/ECI domain as well as a 3-hydroxybutyryl-CoA epimerase domain [2].

The other proteins structurally related to HCDH are:

- Bacterial 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) which reduces 3-hydroxybutanoyl-CoA to acetoacetyl-CoA [3].
- Eye lens protein lambda-crystallin [4], which is specific to lagomorphes (such as rabbit).

There are two major region of similarities in the sequences of proteins of the HCDH family, the first one located in the N-terminal, corresponds to the NAD-binding site, the second one is located in the center of the sequence. A signature pattern has been derived from this central region.

-Consensus pattern: [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-[PA]-[LIVMFY](2)-x(5)-[LIVMFYCT]-[LIVMFY]-x(2)-[GV]

[1] Birktoff J.J., Holden H.M., Hamlin R., Xuong N.-H., Banaszak L.J.
Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987).

[2] Nakahigashi K., Inokuchi H.
Nucleic Acids Res. 18:4937-4937(1990).

[3] Mullany P., Clayton C.L., Pallen M.J., Slone R., Al-Saleh A.,
Tabaqchali S.
FEMS Microbiol. Lett. 124:61-67(1994).

[4] Mulders J.W.M., Hendriks W., Blankesteyn W.M., Bloemendal H.,

de Jong W.W.

J. Biol. Chem. 263:15462-15466(1988).

5 737. 60s Acidic ribosomal protein

Proteins P1, P2, and P0, components of the eukaryotic ribosome stalk. New structural and functional aspects.

Remacha M, Jimenez-Diaz A, Santos C, Briones E, Zambrano R,

Rodríguez Gabriel MA, Guarinos E, Ballesta JP;

10 Biochem Cell Biol 1995;73:959-968.

This family includes archaeobacterial L12, eukaryotic P0, P1 and P2.

738. 6-phosphogluconate dehydrogenases

15 6-phosphogluconate dehydrogenase (EC 1.1.1.44) (6PGD) catalyzes the third step in the hexose monophosphate shunt, the decarboxylating reduction of 6-phosphogluconate in to ribulose 5-phosphate.

Prokaryotic and eukaryotic 6PGD are proteins of about 470 amino acids whose sequence are highly conserved [1]. A region which has been shown [2], from studies of the sheep 6PGD tertiary structure, to be involved in the binding of 6-phosphogluconate has been selected as a signature pattern.

-Consensus pattern: [LIVM]-x-D-x(2)-[GA]-[NQS]-K-G-T-G-x-W

25 [1] Reizer A., Deutscher J., Saier M.H. Jr., Reizer J.

Mol. Microbiol. 5:1081-1089(1991).

[2] Adams M.J., Archibald I.G., Bugg C.E., Came A., Gover S.,

Helliwell J.R., Pickersgill R.W., White S.W.

30 EMBO J. 2:1009-1014(1983).

739. (7tm 1) G-protein coupled receptors [1 to 4,E1,E2] (also called R7G) are an extensive

group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. The receptors that are currently known to belong to this family are listed below.

- 5-hydroxytryptamine (serotonin) 1A to 1F, 2A to 2C, 4, 5A, 5B, 6 and 7 [5].
- Acetylcholine, muscarinic-type, M1 to M5.
- Adenosine A1, A2A, A2B and A3 [6].
- Adrenergic alpha-1A to -1C; alpha-2A to -2D; beta-1 to -3 [7].
- Angiotensin II types I and II.
- Bombesin subtypes 3 and 4.
- Bradykinin B1 and B2.
- c3a and C5a anaphylatoxin.
- Cannabinoid CB1 and CB2.
- Chemokines C-C CC-CKR-1 to CC-CKR-8.
- Chemokines C-X-C CXC-CKR-1 to CXC-CKR-4.
- Cholecystokinin-A and cholecystokinin-B/gastrin.
- Dopamine D1 to D5 [8].
- Endothelin ET-a and ET-b [9].
- fMet-Leu-Phe (fMLP) (N-formyl peptide).
- Follicle stimulating hormone (FSH-R) [10].
- Galanin.
- Gastrin-releasing peptide (GRP-R).
- Gonadotropin-releasing hormone (GNRH-R).
- Histamine H1 and H2 (gastric receptor I).
- Lutropin-choriogonadotropic hormone (LSH-R) [10].
- Melanocortin MC1R to MC5R.
- Melatonin.
- Neuromedin B (NMB-R).
- Neuromedin K (NK-3R).
- Neuropeptide Y types 1 to 6.
- Neurotensin (NT-R).
- Octopamine (tyramine), from insects.

- Odorants [11].
- Opioids delta-, kappa- and mu-types [12].
- Oxytocin (OT-R).
- Platelet activating factor (PAF-R).
- 5 - Prostacyclin.
- Prostaglandin D2.
- Prostaglandin E2, EP1 to EP4 subtypes.
- Prostaglandin F2.
- Purinoreceptors (ATP) [13].
- 10 - Somatostatin types 1 to 5.
- Substance-K (NK-2R).
- Substance-P (NK-1R).
- Thrombin.
- Thromboxane A2.
- 15 - Thyrotropin (TSH-R) [10].
- Thyrotropin releasing factor (TRH-R).
- Vasopressin V1a, V1b and V2.
- Visual pigments (opsins and rhodopsin) [14].
- Proto-oncogene mas.
- 20 - A number of orphan receptors (whose ligand is not known) from mammals and birds.
- *Caenorhabditis elegans* putative receptors C06G4.5, C38C10.1, C43C3.2, T27D1.3 and ZC84.4.
- Three putative receptors encoded in the genome of cytomegalovirus: US27,
- 25 US28, and UL33.
- ECRF3, a putative receptor encoded in the genome of herpesvirus saimiri.

The structure of all these receptors is thought to be identical. They have seven hydrophobic regions, each of which most probably spans the membrane.

30 The N-terminus is located on the extracellular side of the membrane and is often glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. Most, but not all of these

receptors, lack a signal peptide. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved acidic-Arg-aromatic triplet is present in the N-terminal extremity of the second cytoplasmic loop [15] and could be implicated in the interaction with G proteins.

To detect this widespread family of proteins, a pattern that contains the conserved triplet and that also spans the major part of the third transmembrane helix has been developed.

-Consensus pattern: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-
[LIVMNQGA]-x(2)-
[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSTH]-x(2)-
[LIVM]

[1] Strosberg A.D.

Eur. J. Biochem. 196:1-10(1991).

[2] Kerlavage A.R.

Curr. Opin. Struct. Biol. 1:394-401(1991).

[3] Probst W.C., Snyder L.A., Schuster D.I., Brosius J., Sealfon S.C.

DNA Cell Biol. 11:1-20(1992).

[4] Savarese T.M., Fraser C.M.

Biochem. J. 283:1-9(1992).

[5] Branchek T.

Curr. Biol. 3:315-317(1993).

[6] Stiles G.L.

J. Biol. Chem. 267:6451-6454(1992).

[7] Friell T., Kobilka B.K., Lefkowitz R.J., Caron M.G.

Trends Neurosci. 11:321-324(1988).

[8] Stevens C.F.

Curr. Biol. 1:20-22(1991).

[9] Sakurai T., Yanagisawa M., Masaki T.

Trends Pharmacol. Sci. 13:103-107(1992).

[10] Salesse R., Remy J.J., Levin J.M., Jallal B., Garnier J.

Biochimie 73:109-120(1991).

[11] Lancet D., Ben-Arie N.

Curr. Biol. 3:668-674(1993).

5 [12] Uhl G.R., Childers S., Pasternak G.

Trends Neurosci. 17:89-93(1994).

[13] Barnard E.A., Burnstock G., Webb T.E.

Trends Pharmacol. Sci. 15:67-70(1994).

[14] Applebury M.L., Hargrave P.A.

10 Vision Res. 26:1881-1895(1986).

[15] Attwood T.K., Eliopoulos E.E., Findlay J.B.C.

Gene 98:153-159(1991).

(7tm 1) Visual pigments (opsins) retinal binding site

15 Visual pigments [1,2] are the light-absorbing molecules that mediate vision.

They consist of an apoprotein, opsin, covalently linked to the chromophore

cis-retinal. Vision is effected through the absorption of a photon by cis-

retinal which is isomerized to trans-retinal. This isomerization leads to a

change of conformation of the protein. Opsins are integral membrane proteins

20 with seven transmembrane regions that belong to family 1 of G-protein coupled receptors.

In vertebrates four different pigments are generally found. Rod cells, which

mediate vision in dim light, contain the pigment rhodopsin. Cone cells, which

25 function in bright light, are responsible for color vision and contain three

or more color pigments (for example, in mammals: red, blue and green).

In Drosophila, the eye is composed of 800 facets or ommatidia. Each

ommatidium contains eight photoreceptor cells (R1-R8): the R1 to R6 cells are

30 outer cells, R7 and R8 inner cells. Each of the three types of cells (R1-R6,

R7 and R8) expresses a specific opsin.

Proteins evolutionary related to opsins include squid retinochrome, also known

as retinal photoisomerase, which converts various isomers of retinal into 11-cis retinal and mammalian retinal pigment epithelium (RPE) RGR [3], a protein that may also act in retinal isomerization.

The attachment site for retinal in the above proteins is a conserved lysine residue in the middle of the seventh transmembrane helix. The pattern that had been developed includes this residue.

-Consensus pattern: [LIVMWAC]-[PGC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-
[STACP]-
x(2)-[DENF]-[AP]-x(2)-[IY]
[K is the retinal binding site]

[1] Applebury M.L., Hargrave P.A.

Vision Res. 26:1881-1895(1986).

[2] Fryxell K.J., Meyerowitz E.M.

J. Mol. Evol. 33:367-378(1991).

[3] Shen D., Jiang M., Hao W., Tao L., Salazar M., Fong H.K.W.

Biochemistry 33:13117-13125(1994).

The following descriptions of protein family functions are not provided by the Pfam or Prosite databases.

740. BAH

BAH domain. Number of members: 65

[1] Medline: 97074677. Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain. Nicolas RH, Goodwin GH; Gene 1996;175:233-240.

[2] Medline: 99198739. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. Callebaut I, Courvalin J-C, Mornon JP; FEBS letts 1999;446:189-193.

741. ELM2.

ELM2 domain. The ELM2 (Egl-27 and MTA1 homology 2) domain is a small domain of
 5 unknown function. Number of members: 10

742. Euk porin. EUKARYOTIC_PORIN The major protein of the outer mitochondrial
 membrane of eukaryotes is a porin that forms a voltage-dependent anion-selective
 10 channel (VDAC) that behaves as a general diffusion pore for small hydrophilic molecules [1
 to 4]. The channel adopts an open conformation at low or zero membrane potential and a
 closed conformation at potentials above 30-40 mV.

This protein contains about 280 amino acids and its sequence is composed of between 12
 to 16 beta-strands that span the mitochondrial outer membrane. Yeast contains two
 15 members of this family (genes POR1 and POR2); vertebrates have at least three members
 (genes VDAC1, VDAC2 and VDAC3) [5].

A conserved region located at the C-terminal part of these proteins was selected as a
 signature pattern.

20 Consensus pattern[YH]-x(2)-D-[SPCAD]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]-[DNSTA]-
 [DNS]-x(4)-[GSTAN]-[LIVMA]-x-[LIVMY]

[1] Benz R. Biochim. Biophys. Acta 1197:167-196(1994).

[2] Manella C.A. Trends Biochem. Sci. 17:315-320(1992).

25 [3] Dihanich M. Experientia 46:146-153(1990).

[4] Forte M., Guy H.R., Mannella C.A. J. Bioenerg. Biomembr. 19:341-350(1987).

[5] Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J. Genomics 36:192-196(1996).

30 743. Glyco hydor 19

Chitinases family 19 signatures

cross-reference(s) CHITINASE_19_1, CHITINASE_19_2

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 19 (also known as classes IA or I and IB or II)

5 are enzymes from plants that function in the defense against fungal and insect pathogens by destroying their chitin-containing cell wall. Class IA/I and IB/II enzymes differ in the presence (IA/I) or absence (IB/II) of a N-terminal chitin-binding domain (see the relevant entry <PDO00025>). The catalytic domain of these enzymes consist of about 220 to 230 amino acid residues.

10 Two highly conserved regions were selected as signature patterns, the first one is located in the N-terminal section and contains one of the six cysteines which are conserved in most, if not all, of these chitinases and which is probably involved in a disulfide bond.

Consensus pattern C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA]

15 Consensus pattern [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM]

[1] Flach J., Pilet P.-E., Jolles P. *Experientia* 48:701-716(1992).

[2] Henrissat B. *Biochem. J.* 280:309-316(1991).

20 744. MBD

Methyl-CpG binding domain

The Methyl-CpG binding domain (MBD) binds to DNA that contains one or more symmetrically methylated CpGs [1]. DNA methylation in animals is associated with

25 alterations in chromatin structure and silencing of gene expression. MBD has negligible non-specific affinity for DNA. In vitro foot-printing with MeCP2 showed the MBD can protect a 12 nucleotide region surrounding a methyl CpG pair [1]. MBDs are found in several Methyl-CpG binding proteins and also DNA demethylase [2]. Number of members: 11

30 [1] Medline: 94232813. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nan X, Meehan RR, Bird A; *Nucleic Acids Res* 1993;21:4886-4892.

[2] Medline: 99158138. A mammalian protein with specific demethylase activity for mCpG DNA. Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M; *Nature* 1999;397:579-583.

745. Peptidase C1

Eukaryotic thiol (cysteine) proteases active sites

- 5 cross-reference(s) THIOI_PROTEASE_CYS; THIOI_PROTEASE_HIS;
THIOI_PROTEASE_ASN

Eukaryotic thiol proteases (EC 3.4.22.-) [1] are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases which are currently known to belong to this family are listed below (references are only provided for recently determined sequences).

10 - Vertebrate lysosomal cathepsins B (EC 3.4.22.1), H (EC 3.4.22.16), L (EC 3.4.22.15), and S (EC 3.4.22.27) [2].

15 - Vertebrate lysosomal dipeptidyl peptidase I (EC 3.4.14.1) (also known as cathepsin C) [2].

- Vertebrate calpains (EC 3.4.22.17). Calpains are intracellular calcium- activated thiol protease that contain both a N-terminal catalytic domain and a C-terminal calcium-binding domain.

- Mammalian cathepsin K, which seems involved in osteoclastic bone resorption [3].

20 - Human cathepsin O [4].

- Bleomycin hydrolase. An enzyme that catalyzes the inactivation of the antitumor drug BLM (a glycopeptide).

- Plant enzymes: barley aleurain (EC 3.4.22.16), EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin (EC 3.4.22.14); papaya latex papain (EC 3.4.22.2),

25 chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30), and proteinase IV (EC 3.4.22.25); pea turgor-responsive protein 15A; pineapple stem bromelain (EC 3.4.22.32); rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A.

- House-dust mites allergens DerP1 and EurM1.

30 - Cathepsin B-like proteinases from the worms *Caenorhabditis elegans* (genes gcp-1, cpr-3, cpr-4, cpr-5 and cpr-6), *Schistosoma mansoni* (antigen SM31) and *Japonica* (antigen SJ31), *Haemonchus contortus* (genes AC-1 and AC-2), and *Ostertagia ostertagi* (CP-1 and CP-3).

598

- Slime mold cysteine proteinases CP1 and CP2.
- Cruzipain from *Trypanosoma cruzi* and *brucei*.
- Throphozoite cysteine proteinase (TCP) from various *Plasmodium* species.
- Proteases from *Leishmania mexicana*, *Theileria annulata* and *Theileria parva*.
- 5 - Baculoviruses cathepsin-like enzyme (v-cath).
- *Drosophila* small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain.
- Yeast thiol protease BLH1/YCP1/LAP3.
- *Caenorhabditis elegans* hypothetical protein C06G4.2, a calpain-like protein.

10

Two bacterial peptidases are also part of this family:

- Aminopeptidase C from *Lactococcus lactis* (gene pepC) [5].
- Thiol protease tpr from *Porphyromonas gingivalis*.

15

Three other proteins are structurally related to this family, but may have lost their proteolytic activity.

- Soybean oil body protein P34. This protein has its active site cysteine replaced by a glycine.

20

- Rat testin, a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine. Rat testin should not be confused with mouse testin which is a LIM-domain protein (see <PDOC00382>).

25

- *Plasmodium falciparum* serine-repeat protein (SERA), the major blood stage antigen.

This protein of 111 Kd possesses a C-terminal thiol-protease-like domain [6], but the active site cysteine is replaced by a serine.

The sequences around the three active site residues are well conserved and can be used as signature patterns.

30

Consensus pattern Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV] [C is the active site residue]

Note the residue in position 4 of the pattern is almost always cysteine; the only exceptions are calpains (Leu), bleomycin hydrolase (Ser) and yeast YCP1 (Ser). Note the residue in position 5 of the pattern is always Gly except in papaya protease IV where it is Glu.

Consensus pattern[LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH]

[H is the active site residue]

Consensus pattern[FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-[LIVMFYG]-x-[LIVMF] [N is the active site residue]

Note these proteins belong to family C1 (papain-type) and C2 (calpains) in the classification of peptidases [7,E1].

[1]Dufour E. Biochimie 70:1335-1342(1988).

[2]Kirschke H., Barrett A.J., Rawlings N.D. Protein Prof. 2:1587-1643(1995).

[3]Shi G.-P., Chapman H.A., Bhairi S.M., Deleeuw C., Reddy V.Y., Weiss S.J. FEBS Lett. 357:129-134(1995).

[4]Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. J. Biol. Chem. 269:27136-27142(1994).

[5]Chapot-Chartier M.P., Nardi M., Chopin M.C., Chopin A., Gripon J.C. Appl. Environ. Microbiol. 59:330-333(1993).

[6]Higgins D.G., McConnell D.J., Sharp P.M. Nature 340:604-604(1989).

[7]Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:461-486(1994).

746. Peptidase M22

Glycoprotease family signature cross-reference(s) GLYCOPROTEASE

Glycoprotease (GCP) (EC 3.4.24.57) [1], or o-sialoglycoprotein endopeptidase, is a metalloprotease secreted by *Pasteurella haemolytica* which specifically cleaves O-sialoglycoproteins such as glycophorin A. The sequence of GCP is highly similar to the following uncharacterized proteins:

- *Escherichia coli* hypothetical protein ygiD (ORF-X).
- *Bacillus subtilis* hypothetical protein ydiE.
- *Mycobacterium leprae* hypothetical protein U229E.
- *Mycobacterium tuberculosis* hypothetical protein MtCY78.10.

- One of the conserved regions contains two conserved histidines. It is possible that this region is involved in coordinating a metal ion such as zinc.

- Note these proteins belong to family M22 in the classification of peptidases [2,E1].

747. SAM. SAM domain (Sterile alpha motif)

- 20 It has been suggested that SAM is an evolutionarily conserved protein binding domain that is involved in the regulation of numerous developmental processes in diverse eukaryotes. The SAM domain can potentially function as a protein interaction module through its ability to homo- and heterooligomerise with other SAM domains. Number of members: 81

- 25 [1]Medline: 96100659 SAM: A novel motif in yeast sterile alpha and Drosophila
polyhomeotic proteins Ponting CP; Prot Sci 1995;4:1928-1930.
[2]Medline: 97160498 SAM as a protein interaction domain involved in developmental
regulation. Shultz J, Ponting CP, Hofmann K, Bork P; Prot Sci 1997;6:249-253.
[3]Medline: 99101382 The crystal structure of an Eph receptor SAM domain reveals a
30 mechanism for modular dimerization. Reference Author: Stapleton D, Balan I, Pawson
T, Sicheri F; Nat Struct Biol 1999;6:44-49.

748. Tyrosinase signatures cross-reference(s) TYROSINASE_1; TYROSINASE_2
 Tyrosinase (EC 1.14.18.1) [1] is a copper monooxygenases that catalyzes the
 hydroxylation of monophenols and the oxidation of o-diphenols to o-quinols.
 This enzyme, found in prokaryotes as well as in eukaryotes, is involved in the
 5 formation of pigments such as melanins and other polyphenolic compounds.

Tyrosinase binds two copper ions (CuA and CuB). Each of the two copper ion has
 been shown [2] to be bound by three conserved histidines residues. The regions
 around these copper-binding ligands are well conserved and also shared by some
 10 hemocyanins, which are copper-containing oxygen carriers from the hemolymph of
 many molluscs and arthropods [3,4].

At least two proteins related to tyrosinase are known to exist in mammals:

- 15 - TRP-1 (TYRP1) [5], which is responsible for the conversion of 5,6-dihydro-
 xyindole-2-carboxylic acid (DHICA) to indole-5,6-quinone-2-carboxylic acid.
- TRP-2 (TYRP2) [6], which is the melanogenic enzyme DOPachrome tautomerase
 (EC 5.3.3.12) that catalyzes the conversion of DOPachrome to DHICA. TRP-2
 differs from tyrosinases and TRP-1 in that it binds two zinc ions instead
 20 of copper [7].

Other proteins that belong to this family are:

- 25 - Plants polyphenol oxidases (PPO) (EC 1.10.3.1) which catalyze the oxidation
 of mono- and o-diphenols to o-diquinones [8].
- *Caenorhabditis elegans* hypothetical protein C02C2.1.

Two signature patterns for tyrosinase and related proteins have been derived
 The first one contains two of the histidines that bind CuA, and is located in
 30 the N-terminal section of tyrosinase. The second pattern contains a histidine
 that binds CuB, that pattern is located in the central section of the enzyme.

Consensus pattern H-x(4,5)-F-[LIVMFPT]-x-[FW]-H-R-x(2)-[LM]-x(3)-E

[The two H's are copper ligands]

Consensus pattern D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D [H is a copper ligand]

- 5 [1]Lerch K. Prog. Clin. Biol. Res. 256:85-98(1988).
- [2]Jackman M.P., Hajnal A., Lerch K. Biochem. J. 274:707-713(1991).
- [3]Linzen B. Naturwissenschaften 76:206-211(1989).
- [4]Lang W.H., van Holde K.E. Proc. Natl. Acad. Sci. U.S.A. 88:244-248(1991).
- [5]Kobayashi T., Urabe K., Winder A., Jimenez-Cervantes C., Imokawa G., Brewington T., Solano F., Garcia-Borrón J.C., Hearing V.J. EMBO J. 13:5818-5825(1994).
- 10 [6]Jackson I.J., Chambers D.M., Tsukamoto K., Copeland N.G., Gilbert D.J., Jenkins N.A., Hearing V. EMBO J. 11:527-535(1992).
- [7]Solano F., Martínez-Liarte J.H., Jimenez-Cervantes C., Garcia-Borrón J.C., Lozano J.A. Biochem. Biophys. Res. Commun. 204:1243-1250(1994).
- 15 [8]Cary J.W., Lax A.R., Flurkey W.H. Plant Mol. Biol. 20:245-253(1992).

749. (Mur Ligase) Folylpolylglutamate synthase signatures

Folylpolylglutamate synthase (EC 6.3.2.17) (FPGS) [1] is the enzyme of folate metabolism that catalyzes ATP-dependent addition of glutamate moieties to tetrahydrofolate.

Its sequence is moderately conserved between prokaryotes (gene folC) and eukaryotes. We developed two signature patterns based on the conserved regions which are rich in glycine residues and could play a role in the catalytical activity and/or in substrate binding.

Description of pattern(s) and/or profile(s)

Consensus pattern [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)-[LIVM](2)-x(3)-[GSK]

Consensus pattern [LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2)

[1]Shane B., Garrow T., Brenner A., Chen L., Choi Y.J., Hsu J.C., Stover P. Adv. Exp. Med. Biol. 338:629-634(1993).

750. (Peptidase M3) Neutral zinc metallopeptidases, zinc-binding region signature

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily, known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

Family M1

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
- Mammalian aminopeptidase N (EC 3.4.11.2).
- Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
- Yeast aminopeptidase yscII (gene APE2).
- Yeast alanine/arginine aminopeptidase (gene AAP1).
- Yeast hypothetical protein YIL137c.
- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

Family M2

- Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.

Family M3

- Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.
- Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).

604

- Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
- Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).
- Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).
- Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).
- Yeast hypothetical protein YKL134c.

Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of Bacillus.
- Pseudolysin (EC 3.4.24.26) from Pseudomonas aeruginosa (gene lasB).
- Extracellular elastase from Staphylococcus epidermidis.
- Extracellular protease prt1 from Erwinia carotovora.
- Extracellular minor protease smp from Serratia marcescens.
- Vibriolysin (EC 3.4.24.25) from various species of Vibrio.
- Protease prtA from Listeria monocytogenes.
- Extracellular proteinase proA from Legionella pneumophila.

Family M5

- Mycolysin (EC 3.4.24.31) from Streptomyces cacaoi.

Family M6

- Immune inhibitor A from Bacillus thuringiensis (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

Family M7

- Streptomyces extracellular small neutral proteases

Family M8

- Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of Leishmania.

Family M9

- Microbial collagenase (EC 3.4.24.3) from *Clostridium perfringens* and *Vibrio alginolyticus*.

5 Family M10A

- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from *Serratia*.
- Alkaline metalloproteinase from *Pseudomonas aeruginosa* (gene *aprA*).
- Secreted proteases A, B, C and G from *Erwinia chrysanthemi*.
- Yeast hypothetical protein YIL108w.

10

Family M10B

- Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).

15

- Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.

20

- Soybean metalloendoproteinase 1.

Family M11

- *Chlamydomonas reinhardtii* gamete lytic enzyme (GLE).

25 Family M12A

- Astacin (EC 3.4.24.21), a crayfish endoprotease.

- Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border metalloendopeptidase.

30

- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The *Drosophila* homolog of BMP-1 is the dorsal-ventral patterning protein *tolloid*.

- Blastula protease 10 (BP10) from *Paracentrotus lividus* and the related protein SpAN from *Strongylocentrotus purpuratus*.

- *Caenorhabditis elegans* protein toh-2.
- *Caenorhabditis elegans* hypothetical protein F42A10.8.
- Choriolytins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish *Oryzias latipes*. These proteases participate in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.

Family M12B

- Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimereylisin I (EC 3.4.25.52) and II (EC 3.4.25.53).
- Mouse cell surface antigen MS2.

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).
- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.
- Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.
- Peptidase O from *Lactococcus lactis* (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

Family M30

- *Staphylococcus hyicus* neutral metalloprotease.

Family M32

607

- Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from *Thermus aquaticus* which is most active at high temperature.

Family M34

- 5 - Lethal factor (LF) from *Bacillus anthracis*, one of the three proteins composing the anthrax toxin.

Family M35

- 10 - Deuterolysin (EC 3.4.24.39) from *Penicillium citrinum* and related proteases from various species of *Aspergillus*.

Family M36

- Extracellular elastolytic metalloproteinases from *Aspergillus*.

- 15 From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the
20 two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Description of pattern(s) and/or profile(s)

- 25 Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ] [The two H's are zinc ligands] [E is the active site residue]

Sequences known to belong to this class detected by the patternALL, except for members of families M5, M7 and M11.

- 30 Other sequence(s) detected in SWISS-PROT55; including *Neurospora crassa* conidiation-specific protein 13 which could be a zinc-protease.

[1] Jongeneel C.V., Bouvier J., Bairoch A.
FEBS Lett. 242:211-214(1989).

[3]Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B., Stoecker W.

[4]Rawlings N.D., Barrett A.J.

[5]Woessner J. Jr.

10 [6]Hite L.A., Fox J.W., Bjarnason J.B.

Trends Biochem. Sci. 18:324-327(1993).

Trends Cell Biol. 4:179-185(1994).

tRNA pseudouridine synthase is involved in the formation of pseudouridine at the anticodon stem and loop of transfer-RNAs. Pseudouridine is an isomer of uridine (5-(beta-D-ribofuranosyl) uracil), and is the most abundant modified nucleoside found in all cellular RNAs. The TruA-like proteins also exhibit a conserved sequence with a strictly conserved aspartic acid, likely involved in catalysis. Number of members: 25

25 cerevisiae contains one atom of zinc essential for its native conformation and tRNA

recognition. Arluison V, Hountondji C, Robert B, Grosjean H; *Biochemistry* 1998;37:7268-7276.

EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase) (EC 2.5.1.19) catalyzes the sixth step in the biosynthesis from chorismate of the aromatic amino acids (the shikimate pathway) in bacteria (gene *aroA*), plants and fungi (where it is part of a multifunctional

enzyme which catalyzes five consecutive steps in this pathway) [1]. EPSP synthase has been extensively studied as it is the target of the potent herbicide glyphosate which inhibits the enzyme.

The sequence of EPSP from various biological sources shows that the structure of the enzyme has been well conserved throughout evolution. Two conserved regions were selected as signature patterns. The first pattern corresponds to a region that is part of the active site and which is also important for the resistance to glyphosate [2]. The second pattern is located in the C-terminal part of the protein and contains a conserved lysine which seems to be important for the activity of the enzyme.

Description of pattern(s) and/or profile(s)

Consensus pattern[LIVM]-x(2)-[GN]-N-[SA]-G-T-[STA]-x-R-x-[LIVMY]-x-[GSTA]

Consensus pattern[KR]-x-[KH]-E-[CST]-[DNE]-R-[LIVM]-x-[STA]-[LIVMC]-x(2)-[EN]-[LIVMF]-x-[KRA]-[LIVMF]-G

[1]Stallings W.C., Abdel-Megid S.S., Lim L.W., Shieh H.-S., Dayringer H.E., Leimgruber N.K., Stegeman R.A., Anderson K.S., Sikorski J.A., Padgett S.R., Kishore G.M. Proc.

Natl. Acad. Sci. U.S.A. 88:5046-5050(1991).

[2]Padgett S.R., Re D.B., Gaser C.S., Eicholtz D.A., Frazier R.B., Hironaka C.M., Levine E.B., Shah D.M., Fraley R.T., Kishore G.M. J. Biol. Chem. 266:22364-22369(1991).

753. Glyco_hydro_18

Glycosyl hydrolases family 18. Number of members: 173

[1]Medline: 95219379. Crystal structure of a bacterial chitinase at 2.3 Å resolution. Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS, Vorgias CE; Structure 1994;2:1169-1180.

754. Esterase

Putative esterase

This family contains Esterase D Swiss:P10768. However it is not clear if all members of the family have the same function. This family is possibly related to the COesterase family.

Number of members: 36

755. (HMA) Heavy-metal-associated domain

A conserved domain of about 30 amino acid residues has been found [1] in a number of proteins that transport or detoxify heavy metals. This domain contains two conserved cysteines that could be involved in the binding of these metals. The domain has been termed Heavy-Metal-Associated (HMA). It has been found in:

- A variety of cation transport ATPases (E1-E2 ATPases) (see <PDOC00139>). The human copper ATPases ATP7A and ATP7B which are respectively involved in Menke's and Wilson's diseases. ATP7A and ATP7B both contain 6 tandem copies of the HMA domain. The copper ATPases CCC2 from budding yeast, copA from *Enterococcus faecalis* and synA from *Synechococcus* contain one copy of the HMA domain. The cadmium ATPases cadA from *Bacillus firmus* and from plasmid p1258 from *Staphylococcus aureus* also contain a single HMA domain, while a chromosomal *Staphylococcus aureus* cadA contains two copies. Other, less characterized ATPases that contain the HMA domain are: fixI from *Rhizobium meliloti*, pacS from *Synechococcus* strain PCC 7942), *Mycobacterium leprae* ctpA and ctpB and *Escherichia coli* hypothetical protein yhhO. In all these ATPases the HMA domain(s) are located in the N-terminal section.
- Mercuric reductase (EC 1.16.1.1) (gene merA) which is generally encoded by plasmids carried by mercury-resistant Gram-negative bacteria. Mercuric reductase is a class-1 pyridine nucleotide-disulphide oxidoreductase (see <PDOC00073>). There is generally one HMA domain (with the exception of a chromosomal merA from *Bacillus* strain RC607 which has two) in the N-terminal part of merA.
- Mercuric transport protein periplasmic component (gene merP), also encoded by plasmids carried by mercury-resistant Gram-negative bacteria. It seems to be a mercury scavenger that specifically binds to one Hg(2+) ion and which passes it to the mercuric reductase via the merT protein. The N-terminal half of merP is a HMA domain.
- *Helicobacter pylori* copper-binding protein copP.

- Yeast protein ATX1 [2], which could act in the transport and/or partitioning of copper.

The consensus pattern for HMA spans the complete domain.

5

Description of pattern(s) and/or profile(s)

Consensus pattern[LIVN]-x(2)-[LIVMF]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-[IVA]-x-[LVFYS] [The two C's probably bind metals]

10 [1]Bull P.C., Cox D.W. Trends Genet. 10:246-252(1994).

[2]Lin S.-J., Culotta V.L. Proc. Natl. Acad. Sci. U.S.A. 92:3784-3788(1995).

756. (Peptidase M10) Matrixins cysteine switch

PROSITE cross-reference(s): CYSTEINE_SWITCH

15 Mammalian extracellular matrix metalloproteinases (EC 3.4.24.-), also known as matrixins [1] (see <PDOC00129>), are zinc-dependent enzymes. They are secreted by cells in an inactive form (zymogen) that differs from the mature enzyme by the presence of an N-terminal propeptide. A highly conserved octapeptide is found two residues downstream of the C-terminal end of the propeptide. This region has been shown to be involved in
20 autoinhibition of matrixins [2,3]; a cysteine within the octapeptide chelates the active site zinc ion, thus inhibiting the enzyme. This region has been called the 'cysteine switch' or 'autoinhibitor region'.

A cysteine switch has been found in the following zinc proteases:

25 - MMP-1 (EC 3.4.24.7) (interstitial collagenase).

- MMP-2 (EC 3.4.24.24) (72 Kd gelatinase).

- MMP-3 (EC 3.4.24.17) (stromelysin-1).

- MMP-7 (EC 3.4.24.23) (matrilysin).

- MMP-8 (EC 3.4.24.34) (neutrophil collagenase).

30 - MMP-9 (EC 3.4.24.35) (92 Kd gelatinase).

- MMP-10 (EC 3.4.24.22) (stromelysin-2).

- MMP-11 (EC 3.4.24.-) (stromelysin-3).

- MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).

- MMP-13 (EC 3.4.24.-) (collagenase 3).
- MMP-14 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 1).
- MMP-15 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 2).
- MMP-16 (EC 3.4.24.-) (membrane-type matrix metalloproteinase 3).
- 5 - Sea urchin hatching enzyme (EC 3.4.24.12) (envelysin) [4].
- *Chlamydomonas reinhardtii* gamete lytic enzyme (GLE) [5].

Description of pattern(s) and/or profile(s)

Consensus patternP-R-C-[GN]-x-P-[DR]-[LIVSAPKQ] [C chelates the zinc ion]

10

[1]Woessner J. Jr. FASEB J. 5:2145-2154(1991).

[2]Sanchez-Lopez R., Nicholson R., Gesnel M.C., Matrisian L.M., Breathnach R. J. Biol. Chem. 263:11892-11899(1988).

15

[3]Park A.J., Matrisian L.M., Kells A.F., Pearson R., Yuan Z., Navre M. J. Biol. Chem. 266:1584-1590(1991).

[4]Lepage T., Gache C. EMBO J. 9:3003-3012(1990).

[5]Kinoshita T., Fukuzawa H., Shimada T., Saito T., Matsuda Y. Proc. Natl. Acad. Sci. U.S.A. 89:4693-4697(1992).

20

757. (Peptidase S8) Serine proteases, subtilase family, active sites

PROSITE cross-reference(s): PS00136; SUBTILASE_ASP, PS00137; SUBTILASE_HIS, PS00138; SUBTILASE_SER

25

Subtilases [1,2] are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases and can be used as signatures specific to that category of proteases.

30

The subtilase family currently includes the following proteases:

- Subtilisins (EC 3.4.21.62), these alkaline proteases from various *Bacillus* species have been the target of numerous studies in the past thirty years.
- Alkaline elastase YaB from *Bacillus* sp. (gene ale).

- C5A peptidase from *Streptococcus pyogenes* (gene scpA).
 - Cell envelope-located proteases PI, PII, and PIII from *Lactococcus lactis*.
 - Extracellular serine protease from *Serratia marcescens*.
 - Extracellular protease from *Xanthomonas campestris*.
 - Intracellular serine protease (ISP) from various *Bacillus*.
 - Minor extracellular serine protease epr from *Bacillus subtilis* (gene epr).
 - Minor extracellular serine protease vpr from *Bacillus subtilis* (gene vpr).
 - Nisin leader peptide processing protease nisP from *Lactococcus lactis*.
 - Serotype-specific antigene 1 from *Pasteurella haemolytica* (gene ssal).
 - Thermitase (EC 3.4.21.66) from *Thermoactinomyces vulgaris*.
 - Calcium-dependent protease from *Anabaena variabilis* (gene prcA).
 - Halolysin from halophilic bacteria sp. 172p1 (gene hly).
 - Alkaline extracellular protease (AEP) from *Yarrowia lipolytica* (gene xpr2).
 - Alkaline proteinase from *Cephalosporium acremonium* (gene alp).
 - Cerevisin (EC 3.4.21.48) (vacuolar protease B) from yeast (gene PRB1).
 - Cuticle-degrading protease (pr1) from *Metarhizium anisopliae*.
 - KEX-1 protease from *Kluyveromyces lactis*.
 - Kexin (EC 3.4.21.61) from yeast (gene KEX-2).
 - Oryzin (EC 3.4.21.63) (alkaline proteinase) from *Aspergillus* (gene alp).
 - Proteinase K (EC 3.4.21.64) from *Tritirachium album* (gene proK).
 - Proteinase R from *Tritirachium album* (gene proR).
 - Proteinase T from *Tritirachium album* (gene proT).
 - Subtilisin-like protease III from yeast (gene YSP3).
 - Thermomycolin (EC 3.4.21.65) from *Malbranchea sulfurea*.
 - Furin (EC 3.4.21.85), neuroendocrine convertases 1 to 3 (NEC-1 to -3) and PACE4
- protease from mammals, other vertebrates, and invertebrates. These proteases are involved in the processing of hormone precursors at sites comprised of pairs of basic amino acid residues [3].
- Tripeptidyl-peptidase II (EC 3.4.14.10) (tripeptidyl aminopeptidase) from Human.

- Prestalk-specific proteins tagB and tagC from slime mold [4]. Both proteins consist of two domains: a N-terminal subtilase catalytic domain and a C-terminal ABC transporter domain (see <PDOC00185>).

5 Description of pattern(s) and/or profile(s)

Consensus pattern[STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH] [D is the active site residue]

Consensus patternH-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM] [H is the active site residue]

10 Consensus patternG-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG] [S is the active site residue]

Note if a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100%

Note these proteins belong to family S8 in the classification of peptidases [5,E1].

15

[1]Siezen R.J., de Vos W.M., Leunissen J.A.M., Dijkstra B.W. Protein Eng. 4:719-737(1991).

[2]Siezen R.J. (In) Proceeding subtilisin symposium, Hamburg, (1992).

[3]Barr P.J. Cell 66:1-3(1991).

20

[4]Shaulsky G., Kuspa A., Loomis W.F.; Genes Dev. 9:1111-1122(1995).

[5]Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

758. (SSB) Single-strand binding protein family signatures

25 PROSITE cross-reference(s): PS00735; SSB_1,PS00736; SSB_2

The Escherichia coli single-strand binding protein [1] (gene ssb), also known as the helix-destabilizing protein, is a protein of 177 amino acids. It binds tightly, as a homotetramer, to single-stranded DNA (ss-DNA) and plays an important role in DNA replication, recombination and repair.

30

Closely related variants of SSB are encoded in the genome of a variety of large self-transmissible plasmids. SSB has also been characterized in bacteria such as Proteus mirabilis or Serratia marcescens.

Eukaryotic mitochondrial proteins that bind ss-DNA and are probably involved in mitochondrial DNA replication are structurally and evolutionary related to prokaryotic SSB. Proteins currently known to belong to this subfamily are listed below [2].

- 5 - Mammalian protein Mt-SSB (P16).
- *Xenopus* Mt-SSBs and Mt-SSBr.
- *Drosophila* MtSSB.
- Yeast protein RIM1.

- 10 Two signature patterns have been developed for these proteins. The first is a conserved region in the N-terminal section of the SSB's. The second is a centrally located region which, in *Escherichia coli* SSB, is known to be involved in the binding of DNA.

Description of pattern(s) and/or profile(s)

- 15 Consensus pattern[LIVMF]-[NST]-[KRT]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVM]-[GST]-x-[DET]

Consensus patternT-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR]

[1]Meyer R.R., Laine P.S. Microbiol. Rev. 54:342-380(1990).

- 20 [2]Stroumbakis N.D., Li Z., Tolias P.P. Gene 143:171-177(1994).

759. KDPG and KHG aldolases active site signatures

PROSITE cross-reference(s): PS00159; ALDOLASE_KDPG_KHG_1, PS00160;

ALDOLASE_KDPG_KHG_2

25

4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) (KHG-aldolase) catalyzes the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate. Phospho-2-dehydro-3-deoxygluconate aldolase (EC 4.1.2.14) (KDPG-aldolase) catalyzes the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde 3-phosphate.

30

These two enzymes are structurally and functionally related [1]. They are both homotrimeric proteins of approximately 220 amino-acid residues. They are class I aldolases whose catalytic

- Two signature patterns were developed for these enzymes. The first one contains the active site arginine and the second, the lysine involved in the Schiff-base formation.

Consensus pattern G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R [R is the active site residue]

- 10 Consensus patternG-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G [K is involved in Schiff-base formation]

[1] Vlahos C J., Dekker E.E. J. Biol. Chem. 263:11683-11691(1988).

- 15 760. AP endonucleases family 1 signatures. PROSITE cross-reference(s): PS00726;
AP_NUCLEASE_F1_1, PS00727; AP_NUCLEASE_F1_2, PS00728;
AP_NUCLEASE_F1_3

DNA damaging agents such as the antitumor drugs bleomycin and neocarzinostatin or those that generate oxygen radicals produce a variety of lesions in DNA. Amongst these is base-loss which forms apurinic/apyrimidinic (AP) sites or strand breaks with atypical 3'termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of the phosphodiester backbone. Such endonucleases are also generally capable of removing blocking groups from the 3'terminus of DNA strand breaks.

- AP endonucleases can be classified into two families on the basis of sequence similarity. Family 1 groups the enzymes listed below [1].

- 30
- *Escherichia coli* exonuclease III (EC 3.1.11.2) (gene xthA).
 - *Streptococcus pneumoniae* and *Bacillus subtilis* exonuclease A (gene exoA).
 - Mammalian AP endonuclease 1 (AP1) (EC 4.2.99.18).
 - *Drosophila* recombination repair protein 1 (gene Rrp1).
 - *Arabidopsis thaliana* apurinic endonuclease-redox protein (gene arp).

Except for Rrp1 and arp, these enzymes are proteins of about 300 amino-acid residues. Rrp1 and arp both contain additional and unrelated sequences in their N-terminal section (about 400 residues for Rrp1 and 270 for arp).

Three signature patterns were developed for this family of enzymes. The patterns are based on the most conserved regions. The first pattern contains a glutamate which has been shown [2], in the *Escherichia coli* enzyme to bind a divalent metal ion such as magnesium or manganese

Consensus pattern[APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K [E binds a divalent metal ion]
Consensus patternD-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2)
Consensus patternN-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S

[1] Barzilay G., Hickson I.S. BioEssays 17:713-719(1995).
[2] Mol C.D., Kuo C.-F., Thayer M.M., Cunningham R.P., Tainer J.A. Nature 374:381-386(1995).

761. (ER)Enhancer of rudimentary signature, PROSITE cross-reference(s): PS01290; ER

The *Drosophila* protein 'enhancer of rudimentary' (gene (e(r))) is a small protein of 104 residues whose function is not yet clear. From an evolutionary point of view, it is highly conserved [1] and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein plays a role in the cell cycle.

A conserved region in the central part of the protein was selected as as signature pattern.

Consensus patternY-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S

[1] Gelsthorpe M., Pulumati M., McCallum C., Dang-Vu K., Tsubota S.I. Gene 186:189-195(1997).

762. (ETF alpha) Electron transfer flavoprotein alpha-subunit signature, PROSITE cross-reference(s): PS00696; ETF_ALPHA

The electron transfer flavoprotein (ETF) [1,2] serves as a specific electron acceptor for various mitochondrial dehydrogenases. ETF transfers electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase. ETF is an heterodimer that consist of an alpha and a beta subunit and which bind one molecule of FAD per dimer. A similar system also exists in some bacteria.

The alpha subunit of ETF is a protein of about 32 Kd which is structurally related to the bacterial nitrogen fixation protein fixB which could play a role in a redox process and feed electrons to ferredoxin.

Other related proteins are:

- Escherichia coli hypothetical protein ydiR.
- Escherichia coli hypothetical protein ygcQ.

A highly conserved region which is located in the C-terminal section was selected as a signature pattern for these proteins.

Consensus pattern [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)-[IV]-x-A-[IV]-N

- [1] Finocchiaro G., Ikeda Y., Ito M., Tanaka K. Prog. Clin. Biol. Res. 321:637-652(1990).
- [2] Tsai M.H., Saier M.H. Jr. Res. Microbiol. 146:397-404(1995).

763. (lectin c) C-type lectin domain signature and profile
PROSITE cross-reference(s): PS00615; C_TYPE_LLECTIN_1, PS50041;
C_TYPE_LLECTIN_2

A number of different families of proteins share a conserved domain which was first characterized in some animal lectins and which seem to function as a calcium-dependent

carbohydrate-recognition domain [1,2,3]. This domain, which is known as the C-type lectin domain (CTL) or as the carbohydrate-recognition domain (CRD), consists of about 110 to 130 residues. There are four cysteines which are perfectly conserved and involved in two disulfide bonds. A schematic representation of the CTL domain is shown below.

[illegible]

'C': conserved cysteine involved in a disulfide bond.

'c': optional cysteine involved in a disulfide bond.

'*': position of the pattern.

The categories of proteins, in which the CTL domain has been found, are listed below.

Type-II membrane proteins where the CTL domain is located at the C-terminal extremity of the proteins:

- Asialoglycoprotein receptors (ASGPR) (also known as hepatic lectins) [4]. The ASGPR's mediate the endocytosis of plasma glycoproteins to which the terminal sialic acid residue in their carbohydrate moieties has been removed.

- Low affinity immunoglobulin epsilon Fc receptor (lymphocyte IgE receptor), which plays an essential role in the regulation of IgE production and in the differentiation of B cells.

- Kupffer cell receptor. A receptor with an affinity for galactose and fucose, that could be involved in endocytosis.

- A number of proteins expressed on the surface of natural killer T-cells: NKG2, NKR-P1, YE1/88 (Ly-49), CD69 and on B-cells: CD72, LyB-2. The CTL- domain in these proteins is distantly related to other CTL-domains; it is unclear whether they are likely to bind carbohydrates.

Proteins that consist of an N-terminal collagenous domain followed by a CTL-domain [5], these proteins are sometimes called 'collectins':

- Pulmonary surfactant-associated protein A (SP-A). SP-A is a calcium-dependent protein that binds to surfactant phospholipids and contributes to lower the surface tension at the air-liquid interface in the alveoli of the mammalian lung.
- Pulmonary surfactant-associated protein D (SP-D).
- Conglutinin, a calcium-dependent lectin-like protein which binds to a yeast cell wall extract and to immune complexes through the complement component (iC3b).
- Mannan-binding proteins (MBP) (also known as mannose-binding proteins). MBP's bind mannose and N-acetyl-D-glucosamine in a calcium-dependent manner.
- Bovine collectin-43 (CL-43).

Selectins (or LEC-CAM) [6,7]. Selectins are cell adhesion molecules implicated in the interaction of leukocytes with platelets or vascular endothelium. Structurally, selectins consist of a long extracellular domain, followed by a transmembrane region and a short cytoplasmic domain. The extracellular domain is itself composed of a CTL-domain, followed by an EGF-like domain and a variable number of SCR/Sushi repeats. Known selectins are:

- Lymph node homing receptor (also known as L-selectin, leukocyte adhesion molecule-1, (LAM-1), leu-8, gp90-mel, or LECAM-1)
- Endothelial leukocyte adhesion molecule 1 (ELAM-1, E-selectin or LECAM-2). The ligand recognized by ELAM-1 is sialyl-Lewis x.
- Granule membrane protein 140 (GMP-140, P-selectin, PADGEM, CD62, or LECAM-3). The ligand recognized by GMP-140 is Lewis x.

Large proteoglycans that contain a CTL-domain followed by one copy of a SCR/ Sushi repeat, in their C-terminal section:

- Aggrecan (cartilage-specific proteoglycan core protein). This proteoglycan is a major component of the extracellular matrix of cartilaginous tissues where it has a role in the resistance to compression.

- Brevican.

5 - Neurocan.

- Versican (large fibroblast proteoglycan), a large chondroitin sulfate proteoglycan that may play a role in intercellular signalling.

In addition to the CTL and Sushi domains, these proteins also contain, in their N-terminal domain, an Ig-like V-type region, two or four link domains (see <PDOC00955>) and up to two EGF-like repeats.

Two type-I membrane proteins:

15 - Mannose receptor from macrophages. This protein mediates the endocytosis of glycoproteins by macrophages in several recognition and uptake processes. Its extracellular section consists of a fibronectin type II domain followed by eight tandem repeats of the CTL domain.

20 - 180 Kd secretory phospholipase A2 receptor (PLA2-R). A protein whose structure is highly similar to that of the mannose receptor.

- DEC-205 receptor. This protein is used by dendritic cells and thymic epithelial cells to capture and endocytose diverse carbohydrate-binding antigens and direct them to antigen-processing cellular compartments. DEC-205 extracellular section consists of a fibronectin type II domain followed by ten tandem repeats of the CTL domain.

25 - Silk moth hemocytin, an humoral lectin which is involved in a self-defence mechanism. It is composed of 2 FA58C domains (see <PDOC00988>), a CTL domain, 2 VWFC domains (see <PDOC00928>), and a CTCK (see <PDOC00912>).

30 Various other proteins that uniquely consist of a CTL domain:

- Invertebrate soluble galactose-binding lectins. A category to which belong a humoral lectin from a flesh fly; echinoidin, a lectin from the coelomic

fluid of a sea urchin; BRA-2 and BRA-3, two lectins from the coelomic fluid of a barnacle, a lectin from the tunicate *Polyandrocarpa misakiensis* and a newt oviduct lectin. The physiological importance of these lectins is not yet known but they may play an important role in defense mechanisms.

- Pancreatic stone protein (PSP) (also known as pancreatic thread protein (PTP), or reg), a protein that might act as an inhibitor of spontaneous calcium carbonate precipitation.
- Pancreatitis associated protein (PAP), a protein that might be involved in the control of bacterial proliferation.
- Tetranectin, a plasma protein that binds to plasminogen and to isolated kringle 4.
- Eosinophil granule major basic protein (MBP), a cytotoxic protein.
- A galactose specific lectin from a rattlesnake.
- Two subunits of a coagulation factor IX/factor X-binding protein (IX/X-bp), a snake venom anticoagulant protein which binds with factors IX and X in the presence of calcium.
- Two subunits of a phospholipase A2 inhibitor from the plasma of a snake (PLI-A and PLI-B).
- A lipopolysaccharide-binding protein (LPS-BP) from the hemolymph of a cockroach [8].
- Sea raven antifreeze protein (AFP) [9].

As a signature pattern for this domain, the C-terminal region with its three conserved cysteines was selected.

Consensus pattern C-[LIVMFYATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-C [The three C's are involved in disulfide bonds]

Note all CTL domains have five Trp residues before the second Cys, with the exception of tunicate lectin and cockroach LPS-BP which have Leu.

Note this documentation entry is linked to both a signature pattern

and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

- [1] Drickamer K. J. Biol. Chem. 263:9557-9560(1988).
- [2] Drickamer K. Prog. Nucleic Acid Res. Mol. Biol. 45:207-232(1993).
- [3] Drickamer K. Curr. Opin. Struct. Biol. 3:393-400(1993).
- [4] Spiess M. Biochemistry 29:10009-10018(1990).
- [5] Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A. Science 254:1608-1615(1991).
- [6] Siegelman M. Curr. Biol. 1:125-128(1991).
- [7] Lasky L.A. Science 238:964-969(1992).
- [8] Jomori T., Natori S. J. Biol. Chem. 266:13318-13323(1991).
- [9] Ng N.F.L., Hew C.-L. J. Biol. Chem. 267:16069-16075(1992).

764. (SRCR) Speract receptor repeated domain signature

PROSITE cross-reference(s): PS00420; SPERACT_RECEPTOR,

The receptor for the sea urchin egg peptide speract is a transmembrane glycoprotein of 500 amino acid residues [1]. Structurally it consists of a large extracellular domain of 450 residues, followed by a transmembrane region and a small cytoplasmic domain of 12 amino acids. The extracellular domain contains four repeats of a 115 amino acids domain. There are 17 positions that are perfectly conserved in the four repeats, among them are six cysteines, six glycines, and three glutamates.

Such a domain is also found, once, in the C-terminal section of mammalian macrophage scavenger receptor type I [2], a membrane glycoproteins implicated in the pathologic deposition of cholesterol in arterial walls during atherogenesis.

The signature pattern that was derived spans part of the N-terminal section of the domain and contains 8 of the 17 conserved residues.

Consensus pattern G-x(5)-G-x(2)-E-x(6)-W-G-x(2)-C-x(3)-[FYW]-x(8)-C-x(3)-G

[1] Dangott J.J., Jordan J.E., Bellet R.A., Garbers D.L. Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).

[2] Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G., Jenkins N.A.,

5 Krieger M. Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).

765. Bac_surface_Ag

Bacterial surface antigen

This entry includes the following surface antigens; D15 antigen from H.influenzae, OMA87

10 from P.multocida, OMP85 from N.meningitidis and N.gonorrhoeae. Number of members:

14

[1]Medline: 95255676. The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae. Flack FS, Loosmore S, Chong P, Thomas WR; Gene 1995;156:97-99.

[2] Medline: 96333354. Cloning, sequencing, expression, and protective capacity of the oma87 gene encoding the Pasteurella multocida 87-kilodalton outer membrane antigen. Ruffolo CG, Adler B; Infect Immun 1996;64:3161-3167.

20 766. BRCA1 C Terminus (BRCT) domain

The BRCT domain is found predominantly in proteins involved in cell cycle checkpoint functions responsive to DNA damage. It has been suggested that the Retinoblastoma protein contains a divergent BRCT domain, this has not been included in this family. The BRCT domain of XRCC1 forms a homodimer in the crystal structure Medline:99016060. This suggests that pairs of BRCT domains

25 associate as homo- or heterodimers. Number of members: 131

[1] Medline: 96259550. BRCA1 protein products ...Functional motifs... Koonin EV, Altschul SF, Bork P; Nature Genet 1996;13:266-268.

30 [2] Medline: 97153217. From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair Callebaut I, Mornon JP; Febs lett 1997;400:25-30.

[3] Medline: 97186552. A superfamily of conserved domains in DNA damage responsive cell cycle checkpoint proteins Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV; Faseb J 1997;11:68-76.

[4] Medline: 97402527. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ; Nucleic Acids Res 1997;25:3389-3402.

[5] Medline: 99016060. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. Zhang X, Morera S, Bates PA, Whitehead PC, Coffey AI, Hainbuecher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS;

767. Kappa casein

Kappa-casein is a mammalian milk protein involved in a number of important physiological processes. In the gut, the ingested protein is split into an insoluble peptide (para kappa-casein) and a soluble hydrophilic glycopeptide (caseinomacropeptide). Caseinomacropeptide is responsible for increased efficiency of digestion, prevention of neonate hypersensitivity to ingested proteins, and inhibition of gastric pathogens. Number of members: 56

[1] Medline: 98072500. Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae. Ward TJ, Honeycutt RL, Derr JN; Genetics 1997;147:1863-1872.

768. Chitinases family 18 active site

PROSITE cross-reference(s) CHITINASE_18

Chitinases (EC 3.2.1.14) [1] are enzymes that catalyze the hydrolysis of the beta-1,4-N-acetyl-D-glucosamine linkages in chitin polymers. From the view point of sequence similarity chitinases belong to either family 18 or 19 in the classification of glycosyl hydrolases [2,E1]. Chitinases of family 18 (also known as classes III or V) groups a variety of proteins:

a) Chitinases from:

- Prokaryotes such as *Alteromonas*, *Bacillus*, *Serratia*, *Streptomyces*, etc.
- Plants such as *Arabidopsis*, cucumber, bean, tobacco, etc.
- Fungi such as *Aphanocladium*, *Rhizopus*, *Saccharomyces*, etc.

- Nematode (*Brugia malayi*).
- Insects (*Manduca sexta*).
- Baculoviruses (*Autographa Californica* Nuclear Polyhedrosis virus).

5 b) Other proteins:

- Hevamine, a rubber tree protein with chitinase and lysozyme activities.
- *Kluyveromyces lactis* killer toxin alpha subunit, which acts as a chitinase.
- *Flavobacterium* and *Streptomyces* endo-beta-N-acetylglucosaminidases (EC 3.2.1.96).
- 10 - Mammalian di-N-acetylchitobiase which is involved in the degradation of asparagine-linked glycoproteins.
- Human cartilage glycoprotein Gp-39.
- Jack bean concanavalin B (conB), a protein that has lost its catalytic activity.

15 Site directed mutagenesis experiments [3] and crystallographic data [4,5] have shown that a conserved glutamate is involved in the catalytic mechanism and probably acts as a proton donor. This glutamate is at the extremity of the best conserved region in these proteins.

20 Consensus pattern[LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E [E is the active site residue]

- [1] Flach J., Pilet P.-E., Jolles P. *Experientia* 48:701-716(1992).
- [2] Henrissat B. *Biochem. J.* 280:309-316(1991).
- [3] Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H., Uchida M., Tanaka H. *J. Biol. Chem.* 268:18567-18572(1993).
- 25 [4] Perrakis A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S., Vorgias C.E. *Structure* 2:1169-1180(1994).
- [5] van Scheltinga A.C.T., Kalk K.H., Beintema J.J., Dijkstra B.W. *Structure* 2:1181-1189(1994).

30 769. gag_p17. gag gene protein p17 (matrix protein).

The matrix protein forms an icosahedral shell associated with the inner membrane of the mature immunodeficiency virus. Number of members: 1598

[1] Medline: 95055757. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. Massiah MA, Starich MR, Paschall C, Summers MF, Christensen AM, Sundquist WI; J Mol Biol 1994;244:198-223.

5

770. GDA1/CD39 family of nucleoside phosphatases signature

PROSITE cross-reference(s); GDA1_CD39_NTPASE

A number of nucleoside diphosphate and triphosphate hydrolases as well as some yet uncharacterized proteins have been found to belong to the same family [1, 2]. This family currently consist of:

10

- Yeast guanosine-diphosphatase (EC 3.6.1.42) (GDPase) (gene GDA1). GDA1 is a golgi integral membrane enzyme that catalyzes the hydrolysis of GDP to GMP.

- Potato apyrase (EC 3.6.1.5) (adenosine diphosphatase) (ADPase). Apyrase acts on both ATP and ADP to produce AMP.

15

- Mammalian vascular ATP-diphosphohydrolase (EC 3.6.1.5) (also known as lymphoid cell activation antigen CD39).

- Toxoplasma gondii nucleoside-triphosphatases (EC 3.6.1.15) (NTPase). NTPase hydrolyses various nucleoside triphosphates to produce the corresponding nucleoside mono- and diphosphates. This enzyme is secreted into the invaded host cell into the parasitophorous vacuole, a specialized compartment where the parasite intracellularly resides.

20

- Pea nucleoside-triphosphatases (EC 3.6.1.15) (NTPase).

- Caenorhabditis elegans hypothetical protein C33H5.14.

25

- Caenorhabditis elegans hypothetical protein R07E4.4.

- Yeast chromosome V hypothetical protein YER005w.

The above uncharacterized proteins all seem to be membrane-bound.

30

All these proteins share a number of conserved domains. The best conserved of these domains have been selected. It is located in the central section of the proteins.

Consensus pattern[LIVM]-x-G-x(2)-E-G-x-[FY]-x-[FW]-[LIVA]-[TAG]-x-N-[HY]

[1] Handa M., Guidotti G. Biochem. Biophys. Res. Commun. 218:916-923(1996).

[2] Vasconcelos E.G., Ferreira S.T., de Carvalho T.M.U., de Souza W., Kettlun A.M.,

5 Mancilla M., Valenzuela M.A., Verjovski-Almeida S. J. Biol. Chem. 271:22139-22145(1996).

771. GTP cyclohydrolase I signatures

PROSITE cross-reference(s); GTP_CYCLOHYDROL_1_1, GTP_CYCLOHYDROL_1_2

10 GTP cyclohydrolase I (EC 3.5.4.16) catalyzes the biosynthesis of formic acid and dihydroneopterin triphosphate from GTP. This reaction is the first step in the biosynthesis of tetrahydrofolate in prokaryotes, of tetrahydrobiopterin in vertebrates, and of pteridine-containing pigments in insects.

15 GTP cyclohydrolase I is a protein of from 190 to 250 amino acid residues. The comparison of the sequence of the enzyme from bacterial and eukaryotic sources shows that the structure of this enzyme has been extremely well conserved throughout evolution [1].

20 Two conserved regions were selected as signature patterns. The first contains a perfectly conserved tetrapeptide which is part of the GTP-binding pocket [2], the second region also contains conserved residues involved in GTP-binding.

Consensus pattern[DEN]-[LIVM](2)-x(2)-[KRNQ]-[DEN]-[LIVM]-x(3)-[ST]-x-C-E- H-H

Consensus pattern[SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN]

25

[1] Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H. Biochem. Biophys. Res. Commun. 212:705-711(1995).

[2] Nar H., Huber R., Meining W., Schmid C., Weinkauff S., Bacher A. Structure 3:459-466(1995).

30

772. IlvC. Acetohydroxy acid isomeroreductase

Acetohydroxy acid isomeroreductase catalyses the conversion of acetohydroxy acids into dihydroxy valerates. This reaction is the second in the synthetic pathway of the essential branched side chain amino acids valine and isoleucine. Number of members: 29

- 5 [1] Medline: 97361822. The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 Å resolution. Biou V, Dumas R, Cohen-Addad C, Douce R, Job D, Pebay-Peyroula E; EMBO J 1997;16:3405-3415.

10 773. Prokaryotic membrane lipoprotein lipid attachment site

PROSITE cross-reference(s); PROKAR_LIPOPROTEIN

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene lpp).
- Escherichia coli lipoprotein-28 (gene nlpA).
- Escherichia coli lipoprotein-34 (gene nlpB).
- 20 - Escherichia coli lipoprotein nlpC.
- Escherichia coli lipoprotein nlpD.
- Escherichia coli osmotically inducible lipoprotein B (gene osmB).
- Escherichia coli osmotically inducible lipoprotein E (gene osmE).
- Escherichia coli peptidoglycan-associated lipoprotein (gene pal).
- 25 - Escherichia coli rare lipoproteins A and B (genes rplA and rplB).
- Escherichia coli copper homeostasis protein cutF (or nlpE).
- Escherichia coli plasmids traT proteins.
- Escherichia coli Col plasmids lysis proteins.
- A number of Bacillus beta-lactamases.
- 30 - Bacillus subtilis periplasmic oligopeptide-binding protein (gene oppA).
- Borrelia burgdorferi outer surface proteins A and B (genes ospA and ospB).
- Borrelia hermsii variable major protein 21 (gene vmp21) and 7 (gene vmp7).
- Chlamydia trachomatis outer membrane protein 3 (gene omp3).

630

- *Fibrobacter succinogenes* endoglucanase cel-3.

- *Haemophilus influenzae* proteins Pal and Pcp.

- *Klebsiella* pullulanase (gene pulA).

- *Klebsiella* pullulanase secretion protein pulS.

5 - *Mycoplasma hyorhinis* protein p37.

- *Mycoplasma hyorhinis* variant surface antigens A, B, and C (genes vlpABC).

- *Neisseria* outer membrane protein H.8.

- *Pseudomonas aeruginosa* lipopeptide (gene lppL).

- *Pseudomonas solanacearum* endoglucanase egl.

10 - *Rhodospseudomonas viridis* reaction center cytochrome subunit (gene cytC).

- *Rickettsia* 17 Kd antigen.

- *Shigella flexneri* invasion plasmid proteins mxiJ and mxiM.

- *Streptococcus pneumoniae* oligopeptide transport protein A (gene amiA).

- *Treponema pallidum* 34 Kd antigen.

15 - *Treponema pallidum* membrane protein A (gene tmpA).

- *Vibrio harveyi* chitobiase (gene chb).

- *Yersinia* virulence plasmid protein yscJ.

20 - Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, we derived a consensus pattern and a set of rules to identify this type of post-translational modification.

25 Consensus pattern{DERK}(6)-[LIVFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence.

[1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).

30

[2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).

[3] von Heijne G. Protein Eng. 2:531-534(1989).

[4]Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

774. Aminoacyl-transfer RNA synthetases class-II signatures

- 5 PROSITE cross-reference(s); AA_TRNA_LIGASE_II_1; AA_TRNA_LIGASE_II_2
Aminoacyl-tRNA synthetases (EC 6.1.1.-) [1] are a group of enzymes which activate amino acids and transfer them to specific tRNA molecules as the first step in protein biosynthesis. In prokaryotic organisms there are at least twenty different types of aminoacyl-tRNA synthetases, one for each different amino acid. In eukaryotes there are
10 generally two aminoacyl-tRNA synthetases for each different amino acid: one cytosolic form and a mitochondrial form. While all these enzymes have a common function, they are widely diverse in terms of subunit size and of quaternary structure.

- 15 The synthetases specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine are referred to as class-II synthetases [2 to 6] and probably have a common folding pattern in their catalytic domain for the binding of ATP and amino acid which is different to the Rossmann fold observed for the class I synthetases [7].

- 20 Class-II tRNA synthetases do not share a high degree of similarity, however at least three conserved regions are present [2,5,8]. Signature patterns from two of these regions have been derived.

- 25 Consensus pattern[FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE]
Consensus pattern[GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY]

- [1]Schimmel P. Annu. Rev. Biochem. 56:125-158(1987).
[2]Delarue M., Moras D. BioEssays 15:675-687(1993).
30 [3]Schimmel P. Trends Biochem. Sci. 16:1-3(1991).
[4]Nagel G.M., Doolittle R.F. Proc. Natl. Acad. Sci. U.S.A. 88:8121-8125(1991).
[5]Cusack S., Haertlein M., Leberman R. Nucleic Acids Res. 19:3489-3498(1991).
[6]Cusack S. Biochimie 75:1077-1081(1993).

[7]Cusack S., Berthet-Colominas C., Haertlein M., Nassar N., Leberman R. Nature 347:249-255(1990).

[8]Leveque F., Plateau P., Dessen P., Blanquet S. Nucleic Acids Res. 18:305-312(1990).

5 775. X. Trans-activation protein X

This protein is found in hepadnaviruses where it is indispensable for replication. Number of members: 91

776. Thymidylate synthase active site

10 Thymidylate synthase (EC 2.1.1.45) [1,2] catalyzes the reductive methylation of dUMP to dTMP with concomitant conversion of 5,10-methylenetetrahydrofolate to dihydrofolate. Thymidylate synthase plays an essential role in DNA synthesis and is an important target for certain chemotherapeutic drugs.

15 Thymidylate synthase is an enzyme of about 30 to 35 Kd in most species except in protozoan and plants where it exists as a bifunctional enzyme that includes a dihydrofolate reductase domain.

A cysteine residue is involved in the catalytic mechanism (it covalently binds the 5,6-dihydro-dUMP intermediate). The sequence around the active site of this enzyme is conserved from phages to vertebrates.

20 Consensus patternR-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-x-[LV] [C is the active site residue]

[1] Benkovic S.J. Annu. Rev. Biochem. 49:227-251(1980).

25 [2] Ross P., O'Gara F., Condon S. Appl. Environ. Microbiol. 56:2156-2163(1990).

777. Glycosyl hydrolases family 31 signatures

It has been shown [1,2,3,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- 30 - Lysosomal alpha-glucosidase (EC 3.2.1.20) (acid maltase) is a vertebrate glycosidase active at low pH, which hydrolyzes alpha(1->4) and alpha(1->6) linkages in glycogen, maltose, and isomaltose.
- Alpha-glucosidase (EC 3.2.1.20) from the yeast *Candida tsukunbaensis*.

633

- Alpha-glucosidase (EC 3.2.1.20) (gene *malA*) from the archaebacteria *Sulfolobus solfataricus*.

- Intestinal sucrase-isomaltase (EC 3.2.1.48 / EC 3.2.1.10) is a vertebrate membrane-bound, multifunctional enzyme complex which hydrolyzes sucrose, maltose and isomaltose. The sucrase and isomaltase domains of the enzyme are homologous (41% of amino acid identity) and have most probably evolved by duplication.

- Glucoamylase 1 (EC 3.2.1.3) (glucan 1,4-alpha-glucosidase) from various fungal species.

- Yeast hypothetical protein YBR229c.

- Fission yeast hypothetical protein SpAC30D11.01c.

An aspartic acid has been implicated [4] in the catalytic activity of sucrase, isomaltase, and lysosomal alpha-glucosidase. The region around this active residue is highly conserved and can be used as a signature pattern. A second region, which contains two conserved cysteines, has been used as an additional signature pattern.

Consensus pattern [GF]-[LIVMF]-W-x-D-M-[NSA]-E [D is the active site residue]

Consensus pattern G-[AV]-D-[LIVMTA]-C-G-[FY]-x(3)-[ST]-x(3)-L-C-x-R-W-x(2)-[LV]-[GSA]-[SA]-F-x-P-F-x-R-[DN]

[1] Henrissat B. *Biochem. J.* 280:309-316(1991).

[2] Kinsella B.T., Hogan S., Larkin A., Cantwell B.A. *Eur. J. Biochem.* 202:657-664(1991).

[3] Naim H.Y., Niermann T., Kleinhans U., Hollenberg C.P., Strasser A.W.M. *FEBS Lett.* 294:109-112(1991).

[4] Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A., Reuser A.J.J. *J. Biol. Chem.* 266:13507-13512(1991).

778. Urease signatures

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region was selected that contains two histidine that bind one of the nickel ions and the region of the active site histidine.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]

Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]

[1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).

[2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).

[3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

779. Tyrosine specific protein phosphatases signature and profiles

Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) [1 to 5] are enzymes that catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). The currently known PTPases are listed below:

Soluble PTPases.

- PTPN1 (PTP-1B).

- PTPN2 (T-cell PTPase; TC-PTP).

- PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1-like domain (see <PDOC00566>) and could act at junctions between the membrane and cytoskeleton.

- PTPN5 (STEP).

- PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes which contain two copies of the SH2 domain at its N-terminal extremity. The Drosophila protein corkscrew (gene csw) also belongs to this subgroup.

- PTPN7 (LC-PTP; Hematopoietic protein-tyrosine phosphatase; HcPTP).

- PTPN8 (70Z-PEP).

- PTPN9 (MEG2).

- PTPN12 (PTP-G1; PTP-P19).

5 - Yeast PTP1.

- Yeast PTP2 which may be involved in the ubiquitin-mediated protein degradation pathway.

- Fission yeast pyp1 and pyp2 which play a role in inhibiting the onset of mitosis.

- Fission yeast pyp3 which contributes to the dephosphorylation of cdc2.

10 - Yeast CDC14 which may be involved in chromosome segregation.

- Yersinia virulence plasmid PTPases (gene yopH).

- Autographa californica nuclear polyhedrosis virus 19 Kd PTPase.

Dual specificity PTPases.

15 - DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1); which dephosphorylates MAP kinase on both Thr-183 and Tyr-185.

- DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues.

- DUSP3 (VHR).

20 - DUSP4 (HVH2).

- DUSP5 (HVH3).

- DUSP6 (Pyst1; MKP-3).

- DUSP7 (Pyst2; MKP-X).

- Yeast MSG5, a PTPase that dephosphorylates MAP kinase FUS3.

25 - Yeast YVH1.

- Vaccinia virus H1 PTPase; a dual specificity phosphatase.

Receptor PTPases.

30 Structurally, all known receptor PTPases, are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the

PTPase domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.

- 5 In the following table, the domain structure of known receptor PTPases is shown:

	Extracellular	Intracellular
	-----	-----
Ig FN-3 CAH MAM PTPase		

10

Leukocyte common antigen (LCA) (CD45)	0	2	0	0	2
Leukocyte antigen related (LAR)	3	8	0	0	2
Drosophila DLAR	3	9	0	0	2
Drosophila DPTP	2	2	0	0	2
15 PTP-alpha (LRP)	0	0	0	0	2
PTP-beta	0	16	0	0	1
PTP-gamma	0	1	1	0	2
PTP-delta	0	>7	0	0	2
PTP-epsilon	0	0	0	0	2
20 PTP-kappa	1	4	0	1	2
PTP-mu	1	4	0	1	2
PTP-zeta	0	1	1	0	2

25

PTPase domains consist of about 300 amino acids. There are two conserved cysteines, the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.

A signature pattern was derived for PTPase domains centered on the active site cysteine.

30

There are three profiles for PTPases, the first one spans the complete domain and is not specific to any subtype. The second profile is specific to dual-specificity PTPases and the third one to the PTP subfamily.

Consensus pattern [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY] [C is the active site residue]

Notethe M-phase inducer phosphatases (cdc25-type phosphatase) are tyrosine- protein phosphatases that are not structurally related to the above PTPases.

Notethis documentation entry is linked to both a signature pattern and to profiles. As profiles are much more sensitive than the pattern, you should use them if you have access to the necessary software tools to do so.

[1] Fischer E.H., Charbonneau H., Tonks N.K. Science 253:401-406(1991).

[2] Charbonneau H., Tonks N.K. Annu. Rev. Cell Biol. 8:463-493(1992).

[3] Trowbridge I.S. J. Biol. Chem. 266:23517-23520(1991).

[4] Tonks N.K., Charbonneau H. Trends Biochem. Sci. 14:497-500(1989).

[5] Hunter T. Cell 58:1013-1016(1989).

780. Connexins signatures

Gap junctions [1] are specialized regions of the plasma membrane which consist of closely packed pairs of transmembrane channels, the connexons, through which small molecules diffuse from a cell to a neighboring cell. Each connexon is composed of an hexamer of an integral membrane protein which is often referred to as connexin. In a given species there are a number of different, yet structurally related, tissue specific, forms of connexins. The types of connexins which are currently known are listed below.

- Connexin 56 (Cx56).

- Connexin 50 (Cx50) (lens fiber protein MP70).

- Connexin 46 (Cx46) (alpha-3).

- Connexin 45 (Cx45) (alpha-6).

- Connexin 43 (Cx43) (alpha-1).

- Connexin 40 (Cx40) (alpha-5).

- Connexin 38 (Cx38) (alpha-2).

- Connexin 37 (Cx37) (alpha-4).

- Connexin 33 (Cx33) (alpha-7).

- Connexin 32 (Cx32) (beta-1).

- Connexin 31.1 (Cx31.1) (beta-4).

- Connexin 31 (Cx31) (beta-3).

- Connexin 30.3 (Cx30.3) (beta-5).

- Connexin 26 (Cx26) (beta-2).

638

Structurally the connexins consist of a short cytoplasmic N-terminal domain, followed by four transmembrane segments that delimit two extracellular and one cytoplasmic loops; the C-terminal domain is cytoplasmic and its length is variable (from 20 residues in Cx26 to 260 residues in Cx56). The schematic representation of this structure is shown below.

NH2-*** ** *-----COOH

** ** *

** ** * Cytoplasmic

** ** * Membrane

** ** *

** ** * Extracellular

** ** *

** ** *

The sequences of the two extracellular loops are well conserved. In both loops there are three conserved cysteines which are involved in disulfide bonds. A signature patterns from each of these two loop regions has been built.

Consensus pattern C-[DN]-T-x-Q-P-G-C-x(2)-V-C-[FY]-D [The three C's are involved in disulfide bonds] Consensus pattern C-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P [The three C's are involved in disulfide bonds]

[1] Goodenough D.A., Goliger J.A., Paul D.L. Annu. Rev. Biochem. 65:475-502(1996).

781. Gram-positive cocci surface proteins 'anchoring' hexapeptide

Surface proteins from Gram-positive cocci contains a conserved hexapeptide located a few residues downstream of a hydrophobic C-terminal membrane anchor region which is followed by a cluster of basic amino acids [1]. This structure is represented in the following schematic representation:

+-----+-----+

| Variable length extracellular domain |H| Anchor |B|

+-----+-----+-----+

'H': conserved hexapeptide.

'B': cluster of basic residues.

- 5 It has been proposed that this hexapeptide sequence is responsible for a post-translational modification necessary for the proper anchoring of the proteins which bear it, to the cell wall. Proteins known to contain such hexapeptide are listed below:
- Aggregation substance from streptococcus faecalis (asa1).
 - C5a peptidase from Streptococcus pyogenes (scpA).
 - 10 - C protein alpha-antigen from Streptococcus agalactiae (bca).
 - Cell surface antigen I/II (PAC) from Streptococcus mutans.
 - Dextranase from Streptococcus downei (dex).
 - Fibronectin-binding protein from Staphylococcus aureus (fnbA).
 - Fimbrial subunits from Actinomyces naeslundii and viscosus.
 - 15 - IgA binding protein from Streptococcus pyogenes (arp4).
 - IgA binding protein (B antigen) from Streptococcus agalactiae (bag).
 - IgG binding proteins from Streptococci and Staphylococcus aureus.
 - Internalin A from Listeria monocytogenes (inlA).
 - M proteins from streptococci.
 - 20 - Muramidase-released protein from Streptococcus suis (mrp).
 - Nisin leader peptide processing protease from Lactococcus lactis (nisP).
 - Protein A from Staphylococcus aureus.
 - Trypsin-resistant surface T protein from streptococci.
 - Wall-associated protein from Streptococcus mutans (wapA).
 - 25 - Wall-associated serine proteinases from Lactococcus lactis.

Consensus pattern L-P-x-T-G-[STGAVDE]

[1] Schneewind O., Jones K.F., Fischetti V.A. J. Bacteriol. 172:3310-3317(1990).

782. Gamma-glutamyltranspeptidase signature

Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or

water (forming glutamate). GGT plays a key role in the gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. *Pseudomonas cephalosporin acylases* (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region has been used as a signature pattern.

Consensus pattern T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

[1] Tate S.S., Meister A. *Meth. Enzymol.* 113:400-419(1985).

[2] Suzuki H., Kumagai H., Echigo T., Tochikura T. *J. Bacteriol.* 171:5169-5172(1989).

[3] Ishiye M., Niwa M. *Biochim. Biophys. Acta* 1132:233-239(1992).

783. Ferrochelatase signature

Ferrochelatase (EC 4.99.1.1) (protoheme ferro-lyase) [1,2] catalyzes the last step in heme biosynthesis: the chelation of a ferrous ion to proto-porphyrin IX, to form protoheme.

In eukaryotes, ferrochelatase is a mitochondrial protein bound to the inner membrane, whose active site faces the mitochondrial matrix. The mature form of eukaryotic ferrochelatase is composed of about 360 amino acids. In bacteria, ferrochelatase (gene hemH) [3] is a protein of from 310 to 380 amino acids.

The human autosomal dominant disease protoporphyria is due to the reduced activity of ferrochelatase.

The signature pattern for this enzyme is based on a conserved region which contains a histidine residue which could be involved in binding iron.

Consensus pattern [LIVMF](2)-x-[ST]-x-H-[GS]-[LIVM]-P-x(4,5)-[DENQKR]-x-G-[DP]-x(1,2)-Y

[1] Labbe-Bois R. J. Biol. Chem. 265:7278-7283(1990).

[2] Brenner D.A., Frasier F. Proc. Natl. Acad. Sci. U.S.A. 88:849-853(1991).

[3] Miyamoto K., Nakahigashi K., Nishimura K., Inokuchi H. J. Mol. Biol. 219:393-

5 398(1991).

784. Cellulose-binding domain, bacterial type

The microbial degradation of cellulose and xylans requires several types of enzyme such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1].

Structurally, cellulases and xylanases generally consist of a catalytic domain joined to a cellulose-binding domain (CBD) by a short linker sequence rich in proline and/or hydroxy-amino acids.

The CBD of a number of bacterial cellulases has been shown to consist of about 105 amino acid residues [2]. Enzymes known to contain such a domain are:

- Endoglucanase (gene end1) from *Butyrivibrio fibrisolvens*.
- Endoglucanases A (gene cenA) and B (cenB) from *Cellulomonas fimi*.
- Exoglucanases A (gene cbhA) and B (cbhB) from *Cellulomonas fimi*.
- Endoglucanase E-2 (gene celB) from *Thermomonospora fusca*.
- Endoglucanase A (gene celA) from *Microbispora bispora*.
- Endoglucanases A (gene celA), B (celB) and C (celC) from *Pseudomonas fluorescens*.
- Endoglucanase A (gene celA) from *Streptomyces lividans*.
- Exocellobiohydrolase (gene cex) from *Cellulomonas fimi*.
- Xylanases A (gene xynA) and B (xynB) from *Pseudomonas fluorescens*.
- Arabinofuranosidase C (EC 3.2.1.55) (xylanase C) (gene xynC) from *Pseudomonas fluorescens*.
- Chitinase 63 (EC 3.2.1.14) from *Streptomyces plicatus*.
- Chitinase C from *Streptomyces lividans*.

The CBD domain is found either at the N-terminal or at the C-terminal extremity of these enzymes. As it is shown in the following schematic representation, there are two conserved cysteines in this CBD domain - one at each extremity of the domain - which have been shown

'C': conserved cysteine involved in a disulfide bond. '*': position of the pattern.

10 Consensus patternW-N-[STAGR]-[STDN]-[LIVM]-x(2)-[GST]-x-[GST]-x(2)- [LIVMFT]-
[GA]

[1] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[2] Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Protein Seq. Data Anal. 4:349-353(1991).

[3] Gilkes N.R., Claeysens M., Aebersold R., Henrissat B., Meinke A., Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr. Eur. J. Biochem. 202:367-377(1991).

20 785. Amidases signature

It has been shown [1,2,3] that several enzymes from various prokaryotic and eukaryotic organisms which are involved in the hydrolysis of amides (amidases) are evolutionary related. These enzymes are listed below.

- Indoleacetamide hydrolase (EC 3.5.1.1-), a bacterial plasmid-encoded enzyme that catalyzes the hydrolysis of indole-3-acetamide (IAM) into indole-3-acetate (IAA), the second step in the biosynthesis of auxins from tryptophan.

- Acetamidase from *Emericella nidulans* (gene *amdS*), an enzyme which allows acetamide to be used as a sole carbon or nitrogen source.

- Amidase (EC 3.5.1.4) from *Rhodococcus* sp. N-774 and *Brevibacterium* sp. R312 (gene *amdA*). This enzyme hydrolyzes propionamides efficiently, and also at a lower efficiency, acetamide, acrylamide and indoleacetamide.

- Amidase (EC 3.5.1.4) from *Pseudomonas chlororaphis*.

- Glutamyl-tRNA(Gln) amidotransferase subunit A [5].
- Mammalian fatty acid amide hydrolase (gene FAAH) [6].
- A putative amidase from yeast (gene AMD2).
- Mycobacterium tuberculosis putative amidases amiA2, amiB2, amiC and amiD.

10 All these enzymes contain in their central section a highly conserved region rich in glycine, serine, and alanine residues. This region has been used as a signature pattern.

Consensus pattern: G-[GA]-S-[GS]-[GS]-G-x-[GSA]-[GSAVY]-x-[LIVM]-[GSA]-x(6)-[GSAT]-x-[GA]-x-[DE]-x-[GA]-x-S-[LIVM]-R-x-P-[GSAC]

- [1] Mayaux J.-F., Cerbelaud E., Soubrier F., Faucher D., Petre D. J. Bacteriol. 172:6764-6773(1990).
- [2] Hashimoto Y., Nishiyama M., Ikehata O., Horinouchi S., Beppu T. Biochim. Biophys. Acta 1088:225-233(1991).
- [3] Chang T.-H., Abelson J. Nucleic Acids Res. 18:7180-7180(1990).
- [4] Tsuchiya K., Fukuyama S., Kanzaki N., Kanagawa K., Negoro S., Okada H. J. Bacteriol. 171:3187-3191(1989).
- [5] Curnow A.W., Hong K.W., Yuan R., Kim S.I., Martins O., Winkler W., Henkin T.M., Soll D. Proc. Natl. Acad. Sci. U.S.A. 94:11819-11826(1997).
- [6] Cravatt B.F., Giang D.K., Mayfield S.P., Boger D.L., Lerner R.A., Gilula N.B. Nature 384:83-87(1996).

786. Glycosyl hydrolases family 10 active site

The microbial degradation of cellulose and xylans requires several types of enzymes such as endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) (exoglucanases), or xylanases (EC 3.2.1.8) [1,2]. Fungi and bacteria produces a spectrum of cellulolytic enzymes (cellulases) and xylanases which, on the basis of sequence similarities, can be classified into families. One of these families is known as the cellulase family F [3] or as the glycosyl

hydrolases family 10 [4,E1]. The enzymes which are currently known to belong to this family are listed below.

- *Aspergillus awamori* xylanase A (xynA).
- *Bacillus* sp. strain 125 xylanase (xynA).
- 5 - *Bacillus stearothermophilus* xylanase.
- *Butyrivibrio fibrisolvens* xylanases A (xynA) and B (xynB).
- *Caldocellum saccharolyticum* bifunctional endoglucanase/exoglucanase (celB). This protein consists of two domains; it is the N-terminal domain, which has exoglucanase activity, which belongs to this family.
- 10 - *Caldocellum saccharolyticum* xylanase A (xynA).
- *Caldocellum saccharolyticum* ORF4. This hypothetical protein is encoded in the xynABC operon and is probably a xylanase.
- *Cellulomonas fimi* exoglucanase/xylanase (cex).
- *Clostridium stercorarium* thermostable celloxylanase.
- 15 - *Clostridium thermocellum* xylanases Y (xynY) and Z (xynZ).
- *Cryptococcus albidus* xylanase.
- *Penicillium chrysogenum* xylanase (gene xylP).
- *Pseudomonas fluorescens* xylanases A (xynA) and B (xynB).
- *Ruminococcus flavefaciens* bifunctional xylanase XYLA (xynA). This protein consists of
- 20 three domains: a N-terminal xylanase catalytic domain that belongs to family 11 of glycosyl hydrolases; a central domain composed of short repeats of Gln, Asn an Trp, and a C-terminal xylanase catalytic domain that belongs to family 10 of glycosyl hydrolases.
- *Streptomyces lividans* xylanase A (xlnA).
- *Thermoanaerobacter saccharolyticum* endoxylanase A (xynA).
- 25 - *Thermoascus aurantiacus* xylanase.
- *Thermophilic bacterium* Rt8.B4 xylanase (xynA).

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [5], in the exoglucanase from *Cellulomonas fimi*, to be

30 directly involved in glycosidic bond cleavage by acting as a nucleophile. This region has been used as a signature pattern.

Consensus pattern[GTA]-x(2)-[LIVN]-x-[IVMF]-[ST]-E-[LIY]-[DN]-[LIVMF] [E is the active site residue]

[1] Beguin P. Annu. Rev. Microbiol. 44:219-248(1990).

5 [2] Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C. Jr., Warren R.A.J. Microbiol. Rev. 55:303-315(1991).

[3] Henrissat B., Claeysens M., Tomme P., Lemesle L., Mornon J.-P. Gene 81:83-95(1989).

[4] Henrissat B. Biochem. J. 280:309-316(1991).

10 [5] Tull D., Withers S.G., Gilkes N.R., Kilburn D.G., Warren R.A.J., Aebersold R. J. Biol. Chem. 266:15621-15625(1991).

787. Fructose-bisphosphate aldolase class-II signatures

Fructose-bisphosphate aldolase (EC 4.1.2.13) [1,2] is a glycolytic enzyme that catalyzes the reversible aldol cleavage or condensation of fructose-1,6- bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate. There are two classes of fructose-bisphosphate aldolases with different catalytic mechanisms. Class-II aldolases [2], mainly found in prokaryotes and fungi, are homodimeric enzymes which require a divalent metal ion – generally zinc - for their activity.

20 This family also includes the following proteins:

- Escherichia coli galactitol operon protein gatY which catalyzes the transformation of tagatose 1,6-bisphosphate into glycerone phosphate and D- glyceraldehyde 3-phosphate.
- Escherichia coli N-acetyl galactosamine operon protein agaY which catalyzes the same reaction as that of gatY.

25 As signature patterns for this class of enzyme, two conserved regions were selected. The first pattern is located in the first half of the sequence and contains two histidine residues that have been shown [4] to be involved in binding a zinc ion. The second is located in the C-terminal section and contains clustered acidic residues and glycines.

30 Consensus pattern[FYVMT]-x(1,3)-[LIVMH]-[APN]-[LIVM]-x(1,2)-[LIVM]-H-x-D-H-[GACH] [The two H's are zinc ligands]

Consensus pattern[LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E

- [1] Perham R.N. Biochem. Soc. Trans. 18:185-187(1990).
 [2] Marsh J.J., Lebherz H.G. Trends Biochem. Sci. 17:110-113(1992).
 [3] von der Osten C.H., Barbas C.F. III, Wong C.-H., Sinskey A.J. Mol. Microbiol. 3:1625-1637(1989).
 [4] Berry A., Marshall K.E. FEBS Lett. 318:11-16(1993).

788. Prolyl oligopeptidase family serine active site

The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

- Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (*Flavobacterium meningosepticum* and *Aeromonas hydrophila*); there is a high degree of sequence conservation between these sequences.

- *Escherichia coli* protease II (EC 3.4.21.83) (oligopeptidase B) (gene *prtB*) which cleaves peptide bonds on the C-terminal side of lysyl and arginyl residues.

- Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.

- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: *STE13*) which is responsible for the proteolytic maturation of the α -factor precursor.

- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: *DAP2*).

- Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

A conserved serine residue has experimentally been shown (in *E. coli* protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-

terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue]

Note these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

- [1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).
- [2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).
- [3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).
- [4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

789. Formate--tetrahydrofolate ligase signatures

Formate--tetrahydrofolate ligase (EC 6.3.4.3) (formyltetrahydrofolate synthetase) (FTHFS) is one of the enzymes participating in the transfer of one-carbon units, an essential element of various biosynthetic pathways. In many of these processes the transfers of one-carbon units are mediated by the coenzyme tetrahydrofolate (THF). Various reactions generate one-carbon derivatives of THF which can be interconverted between different oxidation states by FTHFS, methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5) and methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9).

In eukaryotes the FTHFS activity is expressed by a multifunctional enzyme, C-1-tetrahydrofolate synthase (C1-THF synthase), which also catalyzes the dehydrogenase and cyclohydrolase activities. Two forms of C1-THF synthases are known [1], one is located in the mitochondrial matrix, while the second one is cytoplasmic. In both forms the FTHFS domain consist of about 600 amino acid residues and is located in the C-terminal section of C1-THF synthase. In prokaryotes FTHFS activity is expressed by a monofunctional homotetrameric enzyme of about 560 amino acid residues [2].

The sequence of FTHFS is highly conserved in all forms of the enzyme. As signature patterns, two regions that are almost perfectly conserved were selected. The first one is a

glycine-rich segment located in the N-terminal part of FTHFS and which could be part of an ATP-binding domain [2]. The second pattern is located in the central section of FTHFS.

Consensus patternG-[LIVM]-K-G-G-A-A-G-G-G-Y

5 Consensus patternV-A-T-[IV]-R-A-L-K-x-[HN]-G-G

[1] Shannon K.W., Rabinowitz J.C. J. Biol. Chem. 263:7717-7725(1988).

[2] Lovell C.R., Przybyla A., Ljungdahl L.G. Biochemistry 29:5687-5694(1990).

10 790. Transthyretin signatures

Transthyretin (prealbumin) [1] is a thyroid hormone-binding protein that seems to transport thyroxine (T4) from the bloodstream to the brain. It is a protein of about 130 amino acids that assembles as a homotetramer and forms an internal channel that binds thyroxine. Transthyretin is mainly synthesized in the brain choroid plexus. In humans, variants of the protein are associated with distinct forms of amyloidosis.

The sequence of transthyretin is highly conserved in vertebrates. A number of uncharacterized proteins also belong to this family:

- Escherichia coli hypothetical protein yedX.
- Bacillus subtilis hypothetical protein yunM.
- 20 - Caenorhabditis elegans hypothetical protein R09H10.3.
- Caenorhabditis elegans hypothetical protein ZK697.8.

Two regions were selected as signature patterns. The first located in the N-terminal extremity starts with a lysine known to be involved in binding T4. The second pattern is located in the C-terminal extremity.

Consensus pattern[KH]-[IV]-L-[DN]-x(3)-G-x-P-A-x(2)-[IV]-x-[IV] [The K binds thyroxine]

Consensus patternY-[TH]-[IV]-[AP]-x(2)-L-S-[PQ]-[FYW]-[GS]-[FY]-[QS]

30 [1] Schreiber G., Richardson S.J. Comp. Biochem. Physiol. 116B:137-160(1997).

791. Dihydropteroate synthase signatures

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites. Most microorganisms must synthesize folate de novo because they lack the active transport system of higher vertebrate cells which allows these organisms to use dietary folates. Enzymes that are involved in the biosynthesis of folates are therefore the target of a variety of antimicrobial agents such as trimethoprim or sulfonamides.

Dihydropteroate synthase (EC 2.5.1.15) (DHPS) catalyzes the condensation of 6-hydroxymethyl-7,8-dihydropteridine pyrophosphate to para-aminobenzoic acid to form 7,8-dihydropteroate. This is the second step in the three steps pathway leading from 6-hydroxymethyl-7,8-dihydropterin to 7,8-dihydrofolate. DHPS is the target of sulfonamides which are substrates analog that compete with para-aminobenzoic acid.

Bacterial DHPS (gene *sul* or *folP*) [1] is a protein of about 275 to 315 amino acid residues which is either chromosomally encoded or found on various antibiotic resistance plasmids. In the lower eukaryote *Pneumocystis carinii*, DHPS is the C-terminal domain of a multifunctional folate synthesis enzyme (gene *fas*) [2].

Two signature patterns for DHPS were developed, the first signature is located in the N-terminal section of these enzymes, while the second signature is located in the central section.

Consensus pattern[LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG]

Consensus pattern[GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P

[1] Slock J., Stahly D.P., Han C.-Y., Six E.W., Crawford I.P. J. Bacteriol. 172:7211-7226(1990).

[2] Volpes F., Dyer M., Scaife J.G., Darby G., Stammers D.K., Delves C.J. Gene 112:213-218(1992).

792. Phosphatidylinositol 3- and 4-kinases signatures

Phosphatidylinositol 3-kinase (PI3-kinase) (EC 2.7.1.137) [1] is an enzyme that phosphorylates phosphoinositides on the 3-hydroxyl group of the inositol ring. The exact function of the three products of PI3-kinase - PI-3-P, PI-3,4-P(2) and PI-3,4,5-P(3) - is not yet known, although it is proposed that they function as second messengers in cell signalling. Currently, three forms of PI3-kinase are known:

- The mammalian enzyme which is a heterodimer of a 110 Kd catalytic chain (p110) and an 85 Kd subunit (p85) which allows it to bind to activated tyrosine protein kinases. There are at least two different types of p100 subunits (alpha and beta).

- Yeast TOR1/DRR1 and TOR2/DRR2 [2], PI3-kinases required for cell cycle activation.

5 Both are proteins of about 280 Kd.

- Yeast VPS34 [3], a PI3-kinase involved in vacuolar sorting and segregation. VPS34 is a protein of about 100 Kd.

- Arabidopsis thaliana and soybean VPS34 homologs.

10 Phosphatidylinositol 4-kinase (PI4-kinase) (EC 2.7.1.67) [4] is an enzyme that acts on phosphatidylinositol (PI) in the first committed step in the production of the second messenger inositol-1,4,5,-trisphosphate. Currently the following forms of PI4-kinases are known:

- Human PI4-kinase alpha.

15 - Yeast PIK1, a nuclear protein of 120 Kd.

- Yeast STT4, a protein of 214 Kd.

The PI3- and PI4-kinases share a well conserved domain at their C-terminal section; this domain seems to be distantly related to the catalytic domain of protein kinases [2]. Two signature patterns were developed from the best conserved parts of this domain.

20 Four additional proteins belong to this family:

- Mammalian FKBP-rapamycin associated protein (FRAP) [5], which acts as the target for the cell-cycle arrest and immunosuppressive effects of the FKBP12-rapamycin complex.

25 - Yeast protein ESR1 [6] which is required for cell growth, DNA repair and meiotic recombination.

- Yeast protein TEL1 which is involved in controlling telomere length.

- Yeast hypothetical protein YHR099w, a distantly related member of this family.

- Fission yeast hypothetical protein SpAC22E12.16C.

30

Consensus pattern[LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q

Consensus pattern[GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x-D-R-H-x(2)-N

- [1] Hiles I.D., Otsu M., Volinia S., Fry M.J., Gout I., Dhand R., Panayotou G., Ruiz-Larrea F., Thompson A., Totty N.F., Hsuan J.J., Courtneidge S.A., Parker P.J., Waterfield M.D. Cell 70:419-429(1992).
- 5 [2] Kunz J., Henriquez R., Schneider U., Deuter-Reinhard M., Movva N., Hall M.N. Cell 73:585-596(1993).
- [3] Schu P.V., Takegawa K., Fry M.J., Stack J.H., Waterfield M.D., Emr S.D. Science 260:88-91(1993).
- [4] Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N. EMBO J. 13:2352-2361(1994).
- 10 [5] Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T., Lane W.S., Schreiber S.L. Nature 369:756-758(1994).
- [6] Kato R., Ogawa H. Nucleic Acids Res. 22:3104-3112(1994).

793. FAD-dependent glycerol-3-phosphate dehydrogenase signatures

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In bacteria [1] it is associated with the utilization of glycerol coupled to respiration. In *Escherichia coli*, two isozymes are known: one expressed under anaerobic conditions (gene *glpA*) and one in aerobic conditions (gene *glpD*). In eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2,3].

These enzymes are proteins of about 60 to 70 Kd which contain a probable FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs from the bacterial or yeast proteins by having an EF-hand calcium-binding region (See <PDOC00018>) in its C-terminal extremity.

Two signature patterns were developed. One based on the first half of the FAD-binding domain and one which corresponds to a conserved region in the central part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

- [1] Austin D., Larson T.J. J. Bacteriol. 173:101-107(1991).
 [2] Roennow B., Kielland-Brandt M.C. Yeast 9:1121-1130(1993).
 [3] Brown L.J., McDonald M.J., Lehn D.A., Moran S.M. J. Biol. Chem. 269:14363-14366(1994).

5

794. NOL1/NOP2/sun family signature

The following proteins seems to be evolutionary related:

10

- Mammalian proliferating-cell nucleolar antigen p120 (gene NOL1) which may play a role in the regulation of the cell cycle and the increased nucleolar activity that is associated with the cell proliferation.

- Yeast nucleolar protein NOP2 (or YNA1) which could be involved in nucleolar function during the onset of growth, and in the maintenance of nucleolar structure.

- Yeast hypothetical protein YBL024w.

- Bacterial protein sun (also known as fmu).

15

- Escherichia coli hypothetical protein yebU.

- Mycobacterium tuberculosis hypothetical protein MtCY21B4.24.

- Methanococcus jannaschii hypothetical protein MJ0026.

20

NOL1 is a protein of 855 residues, NOP2 consists of 618 residues, YBL024w of 684, sun is a protein of about 430 to 450 residues and MJ026 has 274 residues. They share a conserved central domain which contains some highly conserved regions. One of these regions was selected as a signature pattern.

Consensus pattern[FV]-D-[KRA]-[LIVMA]-L-x-D-[AV]-P-C-[ST]-[GA]

25

795. moaA / niFB / pqqE family signature

A number of proteins involved in the biosynthesis of metallo cofactors have been shown [1,2] to be evolutionary related. These proteins are:

30

- Bacterial and archebacterial protein moaA, which is involved in the biosynthesis of the molybdenum cofactor (molybdopterin; MPT).

- Arabidopsis thaliana cnx2, a protein involved in molybdopterin biosynthesis and which is highly similar to moaA.

- Bacillus subtilis narA, which seems to be the moaA ortholog in that bacteria.

- Bacterial protein nifB (or fixZ) which is involved in the biosynthesis of the nitrogenase iron-molybdenum cofactor.

- Bacterial protein pqqE which is involved in the biosynthesis of the cofactor pyrrolo-quinoline-quinone (PQQ).

5 - *Pyrococcus furiosus* cmo, a protein involved in the synthesis of a molybdopterin-based tungsten cofactor.

- *Caenorhabditis elegans* hypothetical protein F49E2.1.

10 All these proteins share, in their N-terminal region, a conserved domain that contains three cysteines. In moaA, these cysteines have been shown [1] to be important for the biological activity. They could be involved in the binding of an iron-sulfur cluster.

Consensus pattern[LIV]-x(3)-C-[NP]-[LIVMF]-[QRS]-C-x-[FYM]-C [The three C's are putative Fe-S ligands

[1] Menendez C., Igloi G., Henninger H., Brandsch R. Arch. Microbiol. 164:142-151(1995).

[2] Hoff T., Schnorr K.M., Meyer C., Caboche M. J. Biol. Chem. 270:6100-6107(1995).

796. Forkhead-associated (FHA) domain profile

20 The forkhead-associated (FHA) domain [1,E1] is a putative nuclear signalling domain found in a variety of otherwise unrelated proteins. The FHA domain comprise approximately 55 to 75 amino acids and contains three highly conserved blocks separated by divergent spacer regions. Currently it has been found in the following proteins:

25 - Four transcription factors that also contain a forkhead (FH) domain: mouse myocyte nuclear factor 1 (MNFI), yeast transcription factor FHL1, which probably controls pre-mRNA processing, and yeast FKH1 and FKH2. In those protein the FHA domain is located N-terminal of the DNA-binding FH domain.

30 - Kinase-associated protein phosphatase (KAPP) from *Arabidopsis thaliana*, a protein which specifically interacts with the receptor-type Ser/Thr-kinase RLK5. In KAPP, the FHA domain maps to a region that interacts with the receptor-type protein kinase RLK5 only if the kinase is phosphorylated on serine residues [2].

- Two protein kinases from yeast that are involved in mediating the nuclear response to DNA damage: DUN1 and SPK1/SAD1 [3]. The latter is the only known protein containing two copies of the FHA domain.

- Protein kinase *cds1* from fission yeast contains a FHA domain and might be the ortholog of

5 SPK1.

- Protein kinase MEK1 from yeast, which is involved in meiotic recombination.

- Human nuclear antigen Ki67 which is expressed only in proliferating cells.

- Yeast hypothetical protein YHR115c, which contains a RING-finger C-terminal of the FHA domain.

10 - Yeast hypothetical proteins L8083.1 and 9346.10, which contain an extensive coiled-coil region C-terminal of the FHA domain.

- *Caenorhabditis elegans* hypothetical protein ZK632.2.

- *Caenorhabditis elegans* hypothetical protein C01G6.5.

- FraH from the prokaryote *Anabaena*, which contains a zinc-finger motif N-terminal of the FHA domain.

- An ORF from the bacterium *Streptomyces*, which is on the opposite strand of the protein kinase *pksl*, overlapping the ORF of the kinase.

[1] Hofmann K.O., Bucher P. Trends Biochem. Sci. 20:347-349(1995).

[2] Stone J.M., Collinge M.A., Smith R.D., Horn M.A., Walker J.C. Science 266:793-795(1994).

[3] Navas T.A., Zhou Z., Elledge S.J. Cell 80:29-39(1995).

797. Ald_Xan_dh_C

25 Aldehyde oxidase and xanthine dehydrogenase, C terminus

[1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber R; Medline: 96072968 "Crystal structure of the xanthine oxidase-related aldehyde oxidoreductase from *D. gigas*." Science 1995;270:1170-1176.

30

Number of members: 54

798. Glyco_hydro_38

Glycosyl hydrolases family 38

Glycosyl hydrolases are key enzymes of carbohydrate metabolism.

Number of members: 20

5

[1] Henrissat B; Medline: 98313424; "Glycosidase families" Biochem Soc Trans 1998;26:153-156.

799. HECT

10

HECT-domain (ubiquitin-transferase).

The name HECT comes from Homologous to the E6-AP Carboxyl Terminus.

Number of members: 43

15

[1] Huibregtse JM, Scheffner M, Beaudenon S, Howley PM; Medline: 95223981; "A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase." Proc Natl Acad Sci U S A 1995;92:2563-2567.

20

800. HRDC

HRDC domain

The HRDC (Helicase and RNase D C-terminal) domain has a putative role in nucleic acid binding. Mutations in the HRDC domain cause human disease.

25

Number of members: 19

[1] Morozov V, Mushegian AR, Koonin EV, Bork P; Medline: 98060076; "A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases" Trends Biochem Sci 1997;22:417-418.

30

801. Integrase

Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc

656

binding domain. The central domain is the catalytic domain [1]. The carboxyl terminal domain is a DNA binding domain [2].

Number of members: 581

5

[1] Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; Medline: 95099322. "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases." Science 1994;266:1981-1986.

10

[2] Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM; Medline: 95359147; "Solution structure of the DNA binding domain of HIV-1 integrase." Biochemistry 1995;34:9826-9833

802. lig_chan

Ligand-gated ion channel

15

This family includes the four transmembrane regions of the ionotropic glutamate receptors and NMDA receptors.

Number of members: 128

20

[1] Tong G, Shepherd D, Jahr CE; Medline: 95184014; "Synaptic desensitization of NMDA receptors by calcineurin." Science 1995;267:1510-1512.

803. RhoGAP

RhoGAP domain

25

GTPase activator proteins towards Rho/Rac/Cdc42-like small GTPases.

Number of members: 97

30

[1] Musacchio A, Cantley LC, Harrison SC; Medline: 97121392; "Crystal structure of the breakpoint cluster region-homology domain from phosphoinositide 3-kinase p85 alpha subunit." Proc Natl Acad Sci U S A 1996;93:14373-14378.

[2] Barrett T, Xiao B, Dodson EJ, Dodson G, Ludbrook SB, Nurmahomed K, Gamblin SJ, Musacchio A, Smerdon SJ, Eccleston JF; Medline: 97162209; "The structure of the GTPase-activating domain from p50rhoGAP." *Nature* 1997;385:458-461.

[3] Rittinger K, Walker PA, Eccleston JF, Nurmahomed K, Owen D, Laue E, Gamblin SJ, Smerdon SJ; Medline: 97404320; "Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP." *Nature* 1997;388:693-697.

[4] Boguski MS, McCormick F; Medline: 94081948; "Proteins regulating Ras and its relatives." *Nature* 1993;366:643-654.

804. vwd
von Willebrand factor type D domain

[1] Bork P; Medline: 93327926; "The modular architecture of a new family of growth regulators related to connective tissue growth factor." *FEBS lett* 1993;327:125-130.

Number of members: 92

805. zf-C4_Topoisom
Topoisomerase DNA binding C4 zinc finger

[1] Tse-Dinh YC, Beran-Steed RK; Medline: 89034032; "Escherichia coli DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains." *J Biol Chem* 1988;263:15857-15859.

[2] Ahumada A, Tse-Dinh YC; Medline: 99011409; "The Zn(II) binding motifs of E. coli DNA topoisomerase I is part of a high-affinity DNA binding domain." *Biochem Biophys Res Commun* 1998;251:509-514.

Number of members: 51

806. AIRC
AIR carboxylase

Members of this family catalyse the decarboxylation of 1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR). This family catalyse the sixth step of de novo purine biosynthesis. Some members of this family contain two copies of this domain. Number of members: 35

807. Bromodomain signature and profile

PROSITE cross-reference(s): PS00633; BROMODOMAIN_1, PS50014;
BROMODOMAIN_2

The bromodomain [1,2,3] is a conserved region of about 70 amino acids found in the following proteins:

- Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1). P250 associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle.
- Human RING3, a protein of unknown function encoded in the MHC class II locus.
- Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein.
- Drosophila female sterile homeotic protein (gene fsh), required maternally for proper expression of other homeotic genes involved in pattern formation, such as Ubx.
- Drosophila brahma protein (gene brm), a protein required for the activation of multiple homeotic genes.
- Mammalian homologs of brahma. In human, three brahma-like proteins are known: SNF2a(hBRM), SNF2b, and BRG1.
- Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation
- Human peregrin (or Br140).
- Yeast BDF1 [3], a transcription factor involved in the expression of a broad class of genes including snRNAs.
- Yeast GCN5, a general transcriptional activator operating in concert with certain other DNA-binding transcriptional activators, such as GCN4, HAP2/3/4 or ADA2.
- Yeast NPS1/STH1, involved in G(2) phase control in mitosis.
- Yeast SNF2/SWI2, which is part of a complex with the SNF5, SNF6, SWI3 and ADR6/SWI1 proteins. This SWI-complex is involved in transcriptional activation.
- Yeast SPT7, a transcriptional activator of Ty elements and possibly other genes.

- *Caenorhabditis elegans* protein cbp-1.
- Yeast hypothetical protein YGR056w.
- Yeast hypothetical protein YKR008w.
- Yeast hypothetical protein L9638.1.

5

Some proteins contain a region which, while similar to some extent to a classical bromodomain, diverges from it by either lacking part of the domain or because of an insertion. These proteins are:

- 10 - Mammalian protein HRX (also known as All-1 or MLL), a protein involved in translocations leading to acute leukemias and which possibly acts as a transcriptional regulatory factor. HRX contains a region similar to the C- terminal half of the bromodomain.
- *Caenorhabditis elegans* hypothetical protein ZK783.4. The bromodomain of this protein has a 23 amino-acid insertion.
- 15 - Yeast protein YTA7. This protein contains a region with significant similarity to the C-terminal half of the bromodomain. As it is a member of the AAA family (see <PDOC00572>) it is also in a functionally different context.

20

The above proteins generally contain a single bromodomain, but some of them contain two copies, this is the case of BDF1, CCG1, fsh, RING3, YKR008w and L9638.1.

The exact function of this domain is not yet known but it is thought to be involved in protein-protein interactions and it may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

25

The consensus pattern that has been developed spans a major part of the bromodomain; a more sensitive detection is available through the use of a profile which spans the whole domain.

30

Consensus pattern[STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY]

- In the be
- In dystro

may play a

- In the sl
- In actin-

- intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:
- In the beta chain of spectrin (or fodrin).
 - In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
 - In the slime mold gelation factor (or ABP-120).
 - In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to

intracellular structures [1]. The actin-binding domain of the first 250 residues of the protein. A similar actin-binding terminal region of many different actin-binding proteins

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy, may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that is

intracellular structures [1]. The actin-binding domain of alpha-actinin seems to consist of the first 250 residues of the protein. A similar actin-binding domain has been found in the C-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD). Dystrophin may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to each other.

intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to

- intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:
- In the beta chain of spectrin (or fodrin).
 - In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
 - In the slime mold gelation factor (or ABP-120).
 - In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to

intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to

intracellular structures [1]. The actin-binding domain of the first 250 residues of the protein. A similar actin-binding terminal region of many different actin-binding proteins

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy, may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that is involved in the formation of the cell's internal structure.

intracellular structures [1]. The actin-binding domain of alpha-actinin seems to reside in the first 250 residues of the protein. A similar actin-binding domain has been found in the N-terminal region of many different actin-binding proteins [2,3]:

- In the beta chain of spectrin (or fodrin).
- In dystrophin, the protein defective in Duchenne muscular dystrophy (DMD) and which may play a role in anchoring the cytoskeleton to the plasma membrane.
- In the slime mold gelation factor (or ABP-120).
- In actin-binding protein ABP-280 (or filamin), a protein that link actin filaments to

[1] Schleicher M., Andre E., Harmann A., Noegel A.A. Dev. Genet. 9:521-530(1988).

[2] Matsudaira P. Trends Biochem. Sci. 16:87-92(1991).

[3] Dubreuil R.R. BioEssays 13:219-226(1991).

- 5 809. (COX1) Heme-copper oxidase subunit I, copper B binding region signature
PROSITE cross-reference(s): PS00077; COX1

Heme-copper respiratory oxidases [1] are oligomeric integral membrane protein complexes that catalyze the terminal step in the respiratory chain: they transfer electrons from cytochrome c or a quinol to oxygen. Some terminal
10 oxidases generate a transmembrane proton gradient across the plasma membrane (prokaryotes) or the mitochondrial inner membrane (eukaryotes). The enzyme complex consists of 3-4 subunits (prokaryotes) up to 13 polypeptides (mammals) of which only the catalytic subunit (equivalent to mammalian subunit 1 (CO I)) is found in all heme-copper respiratory oxidases. The presence of a bimetallic
15 center (formed by a high-spin heme and copper B) as well as a low-spin heme, both ligated to six conserved histidine residues near the outer side of four transmembrane spans within CO I is common to all family members [2-4].

In contrary to eukaryotes the respiratory chain of prokaryotes is branched to
20 multiple terminal oxidases. The enzyme complexes vary in heme and copper composition, substrate type and substrate affinity. The different respiratory oxidases allow the cells to customize their respiratory systems according a variety of environmental growth conditions [1].

- 25 Recently also a component of an anaerobic respiratory chain has been found to contain the copper B binding signature of this family: nitric oxide reductase (NOR) exists in denitrifying species of Archae and Eubacteria.

Enzymes that belong to this family are:

30

- Mitochondrial-type cytochrome c oxidase (EC 1.9.3.1) which uses cytochrome c as electron donor. The electrons are transferred via copper A (Cu(A)) and heme a to the bimetallic center of CO I that is formed by a penta-

coordinated heme a and copper B (Cu(B)). Subunit 1 contains 12 transmembrane regions. Cu(B) is said to be ligated to three of the conserved histidine residues within the transmembrane segments 6 and 7.

- Quinol oxidase from prokaryotes that transfers electrons from a quinol to the binuclear center of polypeptide I. This category of enzymes includes Escherichia coli cytochrome O terminal oxidase complex which is a component of the aerobic respiratory chain that predominates when cells are grown at high aeration.

- FixN, the catalytic subunit of a cytochrome c oxidase expressed in nitrogen-fixing bacteroids living in root nodules. The high affinity for oxygen allows oxidative phosphorylation under low oxygen concentrations. A similar enzyme has been found in other purple bacteria.

- Nitric oxide reductase (EC 1.7.99.7) from Pseudomonas stutzeri. NOR reduces nitrate to dinitrogen. It is a heterodimer of norC and the catalytic subunit norB. The latter contains the 6 invariant histidine residues and 12 transmembrane segments [5].

As a signature pattern the copper-binding region was used.

Consensus pattern[YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H [The three H's are copper B ligands]

Note:cytochrome bd complexes do not belong to this family.

[1]

Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B. J. Bacteriol. 176:5587-5600(1994).

[2]

Castresana J., Luebben M., Saraste M., Higgins D.G. EMBO J. 13:2516-2525(1994).

[3]

Capaldi R.A., Malatesta F., Darley-Usmar V.M. Biochim. Biophys. Acta 726:135-148(1983).

[4]

Holm L., Saraste M., Wikstrom M.

EMBO J. 6:2819-2823(1987).

[5]

5 Saraste M., Castresana J.

FEBS Lett. 341:1-4(1994).

810. (dehydrog_molyb) Eukaryotic molybdopterin oxidoreductases signature
PROSITE cross-reference(s): PS00559; MOLYBDOPTERIN_EUK

10

A number of different eukaryotic oxidoreductases that require and bind a
molybdopterin cofactor have been shown [1] to share a few regions of sequence
similarity. These enzymes are:

15

- Xanthine dehydrogenase (EC 1.1.1.204), which catalyzes the oxidation of
xanthine to uric acid with the concomitant reduction of NAD. Structurally,
this enzyme of about 1300 amino acids consists of at least three distinct
domains: an N-terminal 2Fe-2S ferredoxin-like iron-sulfur binding domain
(see <PDOC00175>), a central FAD/NAD-binding domain and a C-terminal Mo-
pterin domain.

20

- Aldehyde oxidase (EC 1.2.3.1), which catalyzes the oxidation aldehydes into
acids. Aldehyde oxidase is highly similar to xanthine dehydrogenase in its
sequence and domain structure.

25

- Nitrate reductase (EC 1.6.6.1), which catalyzes the reduction of nitrate
to nitrite. Structurally, this enzyme of about 900 amino acids consists of
an N-terminal Mo-pterin domain, a central cytochrome b5-type heme-binding
domain (see <PDOC00170>) and a C-terminal FAD/NAD-binding cytochrome
reductase domain.

30

- Sulfite oxidase (EC 1.8.3.1), which catalyzes the oxidation of sulfite to
sulfate. Structurally, this enzyme of about 460 amino acids consists of an
N-terminal cytochrome b5-binding domain followed by a Mo-pterin domain.

There are a few conserved regions in the sequence of the molybdopterin-binding

domain of these enzymes. The pattern used to detect these proteins is based on one of them. It contains a cysteine residue which could be involved in binding the molybdopterin cofactor.

5 Consensus pattern[GA]-x(3)-[KRNOHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-x(2)-[DE]

[1]

Wootton J.C., Nicolson R.E., Cock J.M., Walters D.E., Burke J.F., Doyle
10 W.A., Bray R.C.
Biochim. Biophys. Acta 1057:157-185(1991).

811. (DNA_ligase) ATP-dependent DNA ligase signatures

PROSITE cross-reference(s): PS00697; DNA_LIGASE_A1, PS00333; DNA_LIGASE_A2

15 DNA ligase (polydeoxyribonucleotide synthase) is the enzyme that joins two DNA fragments by catalyzing the formation of an internucleotide ester bond between phosphate and deoxyribose. It is active during DNA replication, DNA repair and DNA recombination. There are two forms of DNA ligase: one requires ATP (EC 6.5.1.1), the other NAD (EC 6.5.1.2).

20 Eukaryotic, archaeobacterial, virus and phage DNA ligases are ATP-dependent. During the first step of the joining reaction, the ligase interacts with ATP to form a covalent enzyme-adenylate intermediate. A conserved lysine residue
25 is the site of adenylation [1,2].

Apart from the active site region, the only conserved region common to all ATP-dependent DNA ligases is found [3] in the C-terminal section and contains a conserved glutamate as well as four positions with conserved basic residues.

30 Signature patterns were developed for both conserved regions.

Consensus pattern[EDQH]-x-K-x-[DN]-G-x-R-[GACIVM] [K is the active site

residue]

Consensus pattern E-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-
[KRH]-x(3,5)-K-[LIVMFY]-K

- 5 Sequences known to belong to this class detected by the pattern ALL, except
for archbacterial DNA ligases.

[1]

Tomkinson A.E., Totty N.F., Ginsburg M., Lindahl T.

- 10 Proc. Natl. Acad. Sci. U.S.A. 88:400-404(1991).

[2]

Lindahl T., Barnes D.E.

Annu. Rev. Biochem. 61:251-281(1992).

[3]

- 15 Kletzin A.

Nucleic Acids Res. 20:5389-5396(1992).

812. (FAD_Gly3P_dh) FAD-dependent glycerol-3-phosphate dehydrogenase signatures
PROSITE cross-reference(s): PS00977; FAD_G3PDH_1, PS00978; FAD_G3PDH_2

20

FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (GPD) catalyzes
the conversion of glycerol-3-phosphate into dihydroxyacetone phosphate. In
bacteria [1] it is associated with the utilization of glycerol coupled to
respiration. In *Escherichia coli*, two isozymes are known: one expressed under
25 anaerobic conditions (gene *glpA*) and one in aerobic conditions (gene *glpD*). In
eukaryotes, a mitochondrial form of GPD participates in the glycerol phosphate
shuttle in conjunction with an NAD-dependent cytoplasmic GPD (EC 1.1.1.8) [2,
3].

- 30 These enzymes are proteins of about 60 to 70 Kd which contain a probable
FAD-binding domain in their N-terminal extremity. The mammalian enzyme differs
from the bacterial or yeast proteins by having an EF-hand calcium-binding
region (See <PDOC00018>) in its C-terminal extremity.

Two signature patterns were developed. One based on the first half of the FAD-binding domain and one which corresponds to a conserved region in the central part of these enzymes.

Consensus pattern[IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G

Consensus patternG-G-K-x(2)-[GSTE]-Y-R-x(2)-A

[1]

Austin D., Larson T.J.

J. Bacteriol. 173:101-107(1991).

[2]

Roennow B., Kielland-Brandt M.C.

Yeast 9:1121-1130(1993).

[3]

Brown L.J., McDonald M.J., Lehn D.A., Moran S.M.

J. Biol. Chem. 269:14363-14366(1994).

813. (Fapy_DNA_glyco) Formamidopyrimidine-DNA glycosylase signature

PROSITE cross-reference(s): PS01242; FPG

Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) [1] (Fapy-DNA glycosylase)

(gene fpg) is a bacterial enzyme involved in DNA repair and which excise oxidized purine bases to release 2,6-diamino-4-hydroxy-5N-methylformamido-pyrimidine (Fapy) and 7,8-dihydro-8-oxoguanine (8-OxoG) residues. In addition to its glycosylase activity, FPG can also nick DNA at apurinic/apyrimidinic sites (AP sites). FPG is a monomeric protein of about 32 Kd which binds and require zinc for its activity.

The binding site for zinc seems to be located in the C-terminal part of the enzyme where fours conserved and essential [2] cysteines are located. A signature pattern was developed based on this region.

667

Consensus pattern C-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C-x(2)-C-Q

[The four C's are putative zinc ligands]

5 [1]

Duwat P., de Oliveira R., Ehrlich S.D., Boiteux S.

Microbiology 141:411-417(1995).

[2]

O'Connor T.E., Graves R.J., Demurcia G., Castaing B., Laval J.

10 J. Biol. Chem. 268:9063-9070(1993).

814. (G_glu_transpept) Gamma-glutamyltranspeptidase signature

PROSITE cross-reference(s): PS00462; G_GLU_TRANSPEPTIDASE

15 Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT) [1] catalyzes the transfer of the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid, a peptide or water (forming glutamate). GGT plays a key role in the gamma-glutamyl cycle, a pathway for the synthesis and degradation of glutathione. In prokaryotes and eukaryotes, it is an enzyme that consists of
20 two polypeptide chains, a heavy and a light subunit, processed from a single chain precursor. The active site of GGT is known to be located in the light subunit.

25 The sequences of mammalian and bacterial GGT show a number of regions of high similarity [2]. Pseudomonas cephalosporin acylases (EC 3.5.1.-) that convert 7-beta-(4-carboxybutanamido)-cephalosporanic acid (GL-7ACA) into 7-aminocephalosporanic acid (7ACA) and glutaric acid are evolutionary related to GGT and also show some GGT activity [3]. Like GGT, these GL-7ACA acylases, are also composed of two subunits.

30

One of the conserved regions correspond to the N-terminal extremity of the mature light chains of these enzymes. This region was used as a signature pattern.

Consensus pattern T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-x(1,2)-[FY]-G

5 [1]

Tate S.S., Meister A.

Meth. Enzymol. 113:400-419(1985).

[2]

Suzuki H., Kumagai H., Echigo T., Tochikura T.

10 J. Bacteriol. 171:5169-5172(1989).

[3]

Ishiye M., Niwa M.

Biochim. Biophys. Acta 1132:233-239(1992).

15 815. G-protein gamma subunit profile

PROSITE cross-reference(s): PS50058; G_PROTEIN_GAMMA

20 Guanine nucleotide-binding proteins (G proteins) [1] act as intermediaries in the transduction of signals generated by transmembrane receptors. G proteins consist of three subunits (alpha, beta, and gamma). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

25 The gamma subunits are small proteins (from 70 to 110 residues) that are bound to the membrane via a isoprenyl group (either a farnesyl or a geranyl-geranyl) covalently linked to their C-terminus. In mammals there are at least 12 different isoforms of gamma subunits.

30 The *Caenorhabditis elegans* protein egl-10, which is a regulator of G-protein signalling, contains a G-protein gamma-like domain.

A profile was developed that spans the complete length of the gamma

subunit.

[1]

Pennington S.R.

5 Protein Prof. 2:16-315(1995).

816. GNS1/SUR4 family signature

PROSITE cross-reference(s): PS01188; GNS1_SUR4

10 The following group of eukaryotic integral membrane proteins, whose exact function has not yet clearly been established, are evolutionary related [1]:

- Yeast GNS1 [2], a protein involved in synthesis of 1,3-beta-glucan.

15 - Yeast SUR4 (or APA1, SRE1) [3], a protein that could act in a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

- Yeast hypothetical protein YJL196c.

- Caenorhabditis elegans hypothetical protein C40H1.4.

- Caenorhabditis elegans hypothetical protein D2024.3.

20

The proteins have from 290 to 435 amino acid residues. Structurally, they seem to be formed of three sections: a N-terminal region with two transmembrane domains, a central hydrophilic loop and a C-terminal region that contains from one to three transmembrane domains. A conserved region that contains three histidines was
25 selected as a signature pattern. This region is located in the hydrophilic loop.

Consensus pattern L-x-F-L-H-x-Y-H-H

30 [1]

Bairoch A.

Unpublished observations (1996).

[2]

El-Sherbeini M., Clemas J.A.

J. Bacteriol. 177:3227-3234(1995).

[3]

Garcia-Arranz M., Maldonado A.M., Mazon M.J., Portillo F.

5 J. Biol. Chem. 269:18076-18082(1994).

817. Immunoglobulins and major histocompatibility complex proteins signature
PROSITE cross-reference(s): PS00290; IG_MHC

10 The basic structure of immunoglobulin (Ig) [1] molecules is a tetramer of two
light chains and two heavy chains linked by disulfide bonds. There are two
types of light chains: kappa and lambda, each composed of a constant domain
(CL) and a variable domain (VL). There are five types of heavy chains: alpha,
delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and
15 three (in alpha, delta and gamma) or four (in epsilon and mu) constant
domains (CH1 to CH4).

The major histocompatibility complex (MHC) molecules are made of two chains.
In class I [2] the alpha chain is composed of three extracellular domains, a
20 transmembrane region and a cytoplasmic tail. The beta chain (beta-2-
microglobulin) is composed of a single extracellular domain. In class II [3],
both the alpha and the beta chains are composed of two extracellular domains,
a transmembrane region and a cytoplasmic tail.

25 It is known [4,5] that the Ig constant chain domains and a single
extracellular domain in each type of MHC chains are related. These
homologous domains are approximately one hundred amino acids long and
include a conserved intradomain disulfide bond. A small pattern
around the C-terminal cysteine is involved in this disulfide bond which can be used to detect
30 these category of Ig related proteins.

Consensus pattern[FY]-x-C-x-[VA]-x-H-Sequences known to belong to this
class detected by the pattern: Ig heavy chains type Alpha C region : All,

[1]

15

[2]

20

[3]

25

[4]

Nature 282:266-270(1979).

[5]

25

[6]

Nature 331:269-272(1988).

31

818. (IGFBP) Insulin-like growth factor binding proteins signature
PROSITE cross-reference(s): PS00222; IGF_BINDING

The insulin-like growth factors (IGF-I and IGF-II) bind to specific binding

proteins in extracellular fluids with high affinity [1,2,3]. These IGF-binding proteins (IGFBP) prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cells culture. They seem to alter the interaction of IGFs with their cell surface
 5 receptors. There are at least six different IGFBPs and they are structurally related.

The following growth-factor inducible proteins are structurally related to IGFBPs and could function as growth-factor binding proteins [4,5]:

- Mouse protein *cyr61* and its probable chicken homolog, protein CEF-10.
- Human connective tissue growth factor (CTGF) and its mouse homolog, protein FISP-12.
- Vertebrate protein NOV.

As a signature pattern a conserved cysteine-rich region located in the N-terminal section of these proteins is used.

Consensus pattern G-C-[GS]-C-C-x(2)-C-A-x(6)-C

Sequences known to belong to this class detected by the pattern ALL, except for IGFBP-6's.

[1]

Rechler M.M.

Vitam. Horm. 47:1-114(1993).

[2]

Shimasaki S., Ling N.

Prog. Growth Factor Res. 3:243-266(1991).

[3]

Clemmons D.R.

Trends Endocrinol. Metab. 1:412-417(1990).

[4]

Bradham D.M., Igarashi A., Potter R.L., Grotendorst G.R.

J. Cell Biol. 114:1285-1294(1991).

[5]

Maloisel V., Martinerie C., Dambrine G., Plassiart G., Brisac M., Crochet J., Perbal B.

5 Mol. Cell. Biol. 12:10-21(1992).

819. LMWPe : Low molecular weight phosphotyrosine protein phosphatase

Number of members: 34

10 [1]Medline: 94329182, The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. Su XD, Taddei N, Stefani M, Ramponi G, Nordlund P; Nature 1994;370:575-578.

820. (myosin_head) ATP/GTP-binding site motif A (P-loop)

15 PROSITE cross-reference(s): PS00017; ATP_GTP_A

20 From sequence comparisons and crystallographic data analysis it has been shown [1,2,3,4,5,6] that an appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a beta-strand and an alpha-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence [1] or the 'P-loop' [5].

25 There are numerous ATP- or GTP-binding proteins in which the P-loop is found. A number of protein families for which the relevance of the presence of such motif has been noted is listed below:

- ATP synthase alpha and beta subunits (see <PDOC00137>).
- 30 - Myosin heavy chains.
- Kinesin heavy chains and kinesin-like proteins (see <PDOC00343>).
- Dynamins and dynamin-like proteins (see <PDOC00362>).
- Guanylate kinase (see <PDOC00670>).

- Thymidine kinase (see <PDOC00524>).
- Thymidylate kinase (see <PDOC01034>).
- Shikimate kinase (see <PDOC00868>).
- Nitrogenase iron protein family (nifH/frxC) (see <PDOC00580>).
- 5 - ATP-binding proteins involved in 'active transport' (ABC transporters) [7] (see <PDOC00185>).
- DNA and RNA helicases [8,9,10].
- GTP-binding elongation factors (EF-Tu, EF-1alpha, EF-G, EF-2, etc.).
- Ras family of GTP-binding proteins (Ras, Rho, Rab, Ral, Ypt1, SEC4, etc.).
- 10 - Nuclear protein ran (see <PDOC00859>).
- ADP-ribosylation factors family (see <PDOC00781>).
- Bacterial dnaA protein (see <PDOC00771>).
- Bacterial recA protein (see <PDOC00131>).
- Bacterial recF protein (see <PDOC00539>).
- 15 - Guanine nucleotide-binding proteins alpha subunits (Gi, Gs, Gt, G0, etc.).
- DNA mismatch repair proteins mutS family (See <PDOC00388>).
- Bacterial type II secretion system protein E (see <PDOC00567>).

Not all ATP- or GTP-binding proteins are picked-up by this motif. A number of proteins escape detection because the structure of their ATP-binding site is completely different from that of the P-loop. Examples of such proteins are the E1-E2 ATPases or the glycolytic kinases. In other ATP- or GTP-binding proteins the flexible loop exists in a slightly different form; this is the case for tubulins or protein kinases. A special mention must be reserved for

25 adenylate kinase, in which there is a single deviation from the P-loop pattern: in the last position Gly is found instead of Ser or Thr.

Consensus pattern[AG]-x(4)-G-K-[ST]

[1]

Walker J.E., Saraste M., Runswick M.J., Gay N.J.
EMBO J. 1:945-951(1982).

[2]

Moller W., Amons R.

FEBS Lett. 186:1-7(1985).

[3]

Fry D.C., Kuby S.A., Mildvan A.S.

5 Proc. Natl. Acad. Sci. U.S.A. 83:907-911(1986).

[4]

Dever T.E., Glynnias M.J., Merrick W.C.

Proc. Natl. Acad. Sci. U.S.A. 84:1814-1818(1987).

[5]

10 Saraste M., Sibbald P.R., Wittinghofer A.

Trends Biochem. Sci. 15:430-434(1990).

[6]

Koonin E.V.

J. Mol. Biol. 229:1165-1174(1993).

15 [7]

Higgins C.F., Hyde S.C., Mimmack M.M., Gileadi U., Gill D.R., Gallagher M.P.

J. Bioenerg. Biomembr. 22:571-592(1990).

[8]

20 Hodgman T.C.

Nature 333:22-23(1988) and Nature 333:578-578(1988) (Errata).

[9]

Linder P., Lasko P., Ashburner M., Leroy P., Nielsen P.J., Nishi K.,

Schnier J., Slonimski P.P.

25 Nature 337:121-122(1989).

[10]

Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M.

Nucleic Acids Res. 17:4713-4730(1989).

30 821. PE: PE family

This family named after a PE motif near to the amino terminus of the domain. The PE family of proteins all contain an amino-terminal region of about 110 amino acids. The carboxyl terminus of this family are variable and fall into several classes. The largest class of PE

proteins is the highly repetitive PGRS class which have a high glycine content. The function of these proteins is uncertain but it has been suggested that they may be related to antigenic variation of *Mycobacterium tuberculosis* [1]. Number of members: 88

- 5 [1] Medline: 98295987. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al; Nature 1998;393:537-544.

10 822. (RNB) Ribonuclease II family signature

PROSITE cross-reference(s): PS01175; RIBONUCLEASE_II

On the basis of sequence similarities, the following bacterial and eukaryotic proteins seem to form a family:

- *Escherichia coli* and related bacteria ribonuclease II (EC 3.1.13.1) (RNase II) (gene *rnB*) [1]. RNase II is an exonuclease involved in mRNA decay. It degrades mRNA by hydrolyzing single-stranded polyribonucleotides processively in the 3' to 5' direction.
- Bacterial protein *vacB*. In *Shigella flexneri*, *vacB* has been shown to be required for the expression of virulence genes at the posttranscriptional level.
- Yeast protein SSD1 (or SRK1) which is implicated in the control of the cell cycle G1 phase.
- Yeast protein DIS3 [2], which binds to ran (GSP1) and enhances the nucleotide-releasing activity of RCC1 on ran.
- Fission yeast protein *dis3*, which is implicated in mitotic control.
- *Neurospora crassa* *cyt-4*, a mitochondrial protein required for RNA 5' and 3' end processing and splicing.
- Yeast protein MSU1, which is involved in mitochondrial biogenesis.
- *Synechocystis* strain PCC 6803 protein *zam* [3], which control resistance to the carbonic anhydrase inhibitor acetazolamide.

- *Caenorhabditis elegans* hypothetical protein F48E8.6.

The size of these proteins range from 644 residues (rnb) to 1250 (SSD1). While their sequence is highly divergent they share a conserved domain in their C-terminal section [4]. It is possible that this domain plays a role in a putative exonuclease function that would be common to all these proteins. A signature pattern was developed based on the core of this conserved domain.

Consensus pattern[HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-[RQ]-[KR]-[FY]-x-D-x(3)-[HQ]

[1]

Zilhao R., Camelo L., Arraiano C.M.

Mol. Microbiol. 8:43-51(1993).

[2]

Noguchi E., Hayashi N., Azuma Y., Seki T., Nakamura M., Nakashima N.,

Yanagida M., He X., Mueller U., Sazer S., Nishimoto T.

EMBO J. 15:5595-5605(1996).

[3]

Beuf L., Bedu S., Cami B., Joset F.

Plant Mol. Biol. 27:779-788(1995).

[4]

Mian I.S.

Nucleic Acids Res. 25:3187-3195(1997).

823. Src homology 2 (SH2) domain profile

PROSITE cross-reference(s): PS50001; SH2

The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with

high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6].

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets [7].

So far, SH2 domains have been identified in the following proteins:

- Many vertebrate, invertebrate and retroviral cytoplasmic (non-receptor) protein tyrosine kinases. In particular in the Src, Abl, Bkt, Csk and ZAP70 families of kinases.
- Mammalian phosphatidylinositol-specific phospholipase C gamma-1 and -2. Two copies of the SH2 domain are found in those proteins in between the catalytic 'X-' and 'Y-boxes' (see <PDOC50007>).
- Mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit.
- Some vertebrate and invertebrate protein-tyrosine phosphatases.
- Mammalian Ras GTPase-activating protein (GAP).
- Adaptor proteins mediating binding of guanine nucleotide exchange factors to growth factor receptors: vertebrate GRB2, Caenorhabditis elegans sem-5 and Drosophila DRK.
- Mammalian Vav oncoprotein, a guanine-nucleotide exchange factor of the CDC24 family.
- Miscellaneous proteins interacting with vertebrate receptor protein tyrosine kinases: oncoprotein Crk, mammalian cytoplasmic proteins Nck, Shc.
- STAT proteins (signal transducers and activators of transcription).
- Chicken tensin.
- Yeast transcriptional control protein SPT6.

The profile developed to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

[1]

Sadowski I., Stone J.C., Pawson T.
Mol. Cell. Biol. 6:4396-4408(1986).

[2]

5 Russel R.B., Breed J., Barton G.J.
FEBS Lett. 304:15-20(1992).

[3]

Marangere L.E.M., Pawson T.
J. Cell Sci. Suppl. 18:97-104(1994).

10 [4]

Pawson T., Schlessinger J.
Curr. Biol. 3:434-442(1993).

[5]

15 Mayer B.J., Baltimore D.
Trends Cell. Biol. 3:8-13(1993).

[6]

Pawson T.
Nature 373:573-580(1995).

[7]

20 Kuriyan J., Cowburn D.
Curr. Opin. Struct. Biol. 3:828-837(1993).

824. Sulfate transporters signature

PROSITE cross-reference(s): PS01130; SULFATE_TRANSP

25

A number of proteins involved in the transport of sulfate across a membrane as well as some yet uncharacterized proteins have been shown [1,2] to be evolutionary related. These proteins are:

- 30 - Neurospora crassa sulfate permease II (gene *cys-14*).
 - Yeast sulfate permeases (genes *SUL1* and *SUL2*).
 - Rat sulfate anion transporter 1 (*SAT-1*).
 - Mammalian DTDST, a probable sulfate transporter which, in Human, is

680

involved in the genetic disease, diastrophic dysplasia (DTD).

- Sulfate transporters 1, 2 and 3 from the legume *Stylosanthes hamata*.

- Human pendrin (gene PDS), which is involved in a number of hearing loss

genetic diseases.

- Human protein DRA (Down-Regulated in Adenoma).

- Soybean early nodulin 70.

- *Escherichia coli* hypothetical protein ychM.

- *Caenorhabditis elegans* hypothetical protein F41D9.5.

As expected by their transport function, these proteins are highly hydrophobic and seem to contain about 12 transmembrane domains. The best conserved region seems to be located in the second transmembrane region and is used as a signature pattern.

Consensus pattern[PAV]-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVFYA]-[LIVST]-[YI]-x(3)-[GA]-[GST]-S-[KR]

[1]

Sandal N.N., Marcker K.A.

Trends Biochem. Sci. 19:19-19(1994).

[2]

Smith F.W., Hawkesford M.J., Prosser I.M., Clarkson D.T.

Mol. Gen. Genet. 247:709-715(1995).

825. TYA: TYA transposon protein

Ty are yeast transposons. A 5.7kb transcript codes for p3 a fusion protein of TYA and TYB.

The TYA protein is analogous to the gag protein of retroviruses. TYA a is cleaved to form 46kd protein which can form mature virion like particles [1]. Number of members: 59

[1] Medline: 97404699. Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like particles. Palmer KJ, Tichelaar W, Myers N, Burns NR, Butcher SJ, Kingsman AJ, Fuller SD, Saibil HR; J Virol 1997;71:6863-6868.

826. Aldolase_II

Class II Aldolase and Adducin N-terminal domain.

-!- This family includes class II aldolases and adducins which have not been ascribed any enzymatic function. Number of members: 37

References:

[1] Medline: 93294819. The spatial structure of the class II L-fucose-1-phosphate aldolase from *Escherichia coli*. Dreyer MK, Schulz GE; J Mol Biol 1993;231:549-553.

[2] Medline: 96256522. Catalytic mechanism of the metal-dependent fucose aldolase from *Escherichia coli* as derived from the structure. Dreyer MK, Schulz GE; J Mol Biol 1996;259:458-466.

827. CBD_2

-!- Two tryptophan residues are involved in cellulose binding.

-!- Cellulose binding domain found in bacteria. Number of members: 51

References:

[1] Medline: 95284032. Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Harvey TS; Biochemistry 1995;34:6993-7009.

828. P

A unique feature of the eukaryotic subtilisin-like proprotein convertases is the presence of an additional highly conserved sequence of approximately 150 residues (P domain) located immediately downstream of the catalytic domain.

Number of members: 91

References:

[1] Medline: 94252314. A C-terminal domain conserved in precursor processing proteases is required for intramolecular N-terminal maturation of pro-Kex2 protease. Gluschankof P, Fuller RS; EMBO J 1994;13:2280-2288.

[2] Medline: 98225190. Regulatory roles of the P domain of the subtilisin-like prohormone convertases. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF; J Biol Chem 1998;273:11107-11114.

5 829. Uncharacterized protein family UPF0020 signature

PROSITE cross-reference(s): PS01261; UPF0020

The following uncharacterized proteins have been shown [1] to share regions of similarities:

- 10 - Escherichia coli hypothetical protein ycbY and HI0116/15, the corresponding Haemophilus influenzae protein.
- Bacillus subtilis hypothetical protein ypsC.
- Synechocystis strain PCC 6803 hypothetical protein slr0064.
- Methanococcus jannaschii hypothetical proteins MJ0438 and MJ0710.

15

These are hydrophilic proteins of from 40 Kd to about 80 Kd. They can be picked up in the database by the following pattern.

Consensus patternD-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E

20

References:

[1] Bairoch A. Unpublished observations (1997).

830. Uncharacterized protein family UPF0031 signatures

25 PROSITE cross-reference(s): PS01049; UPF0031_1; PS01050; UPF0031_2

The following uncharacterized proteins have been shown [1] to share regions of similarities:

30

- Yeast chromosome XI hypothetical protein YKL151c.
- Caenorhabditis elegans hypothetical protein R107.2.
- Escherichia coli hypothetical protein yjeF.
- Bacillus subtilis hypothetical protein yxkO.
- Helicobacter pylori hypothetical protein HP1363.

683

- Mycobacterium tuberculosis hypothetical protein MtCY77.05c.
- Mycobacterium leprae hypothetical protein B229_C2_201.
- Synechocystis strain PCC 6803 hypothetical protein slh1433.
- Methanococcus jannaschii hypothetical protein MJ1586.

5

These are proteins of about 30 to 40 Kd whose central region is well conserved. They can be picked up in the database by the following patterns.

Consensus pattern[SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT]

10

Consensus pattern[GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM]

831. (ACOX)

Acyl-CoA oxidase

15

This is a family of Acyl-CoA oxidases EC:1.3.3.6. Acyl-coA oxidase converts acyl-CoA into trans-2-enoyl-CoA [1].

Number of members: 39

20

[1] Hayashi H, De Bellis L, Yamaguchi K, Kato A, Hayashi M, Nishimura M; Medline: 98192624. "Molecular characterization of a glyoxysomal long chain acyl-CoA oxidase that is synthesized as a precursor of higher molecular mass in pumpkin." J Biol Chem 1998;273:8301-8307.

25

832. (AICARFT_IMPCHas)

AICARFT/IMPCHase bienzyme

30

This is a family of bifunctional enzymes catalysing the last steps in de novo purine biosynthesis. The bifunctional enzyme is found in both prokaryotes and eukaryotes. The second last step is catalysed by 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase EC:2.1.2.3 (AICARFT), this enzyme catalyses the formylation of AICAR with 10-formyl-tetrahydrofolate to yield FAICAR and tetrahydrofolate [1]. The last step is

catalysed by IMP (Inosine monophosphate) cyclohydrolase EC:3.5.4.10 (IMPHase), cyclizing FAICAR (5-formylaminoimidazole-4-carboxamide ribonucleotide) to IMP [1].

Number of members: 22

[1] Akira T, Komatsu M, Nango R, Tomooka A, Konaka K, Yamauchi M, Kitamura Y, Nomura S, Tsukamoto I; Medline: 97473523 "Molecular cloning and expression of a rat cDNA encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase" [published erratum appears in Gene 1998 Feb 27;208(2):337] Gene 1997;197:289-293.

[2] Rayl EA, Moroson BA, Beardsley GP; Medline: 96147205 "The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping." J Biol Chem 1996;271:2225-2233.

833. (AOX)

Alternative oxidase

The alternative oxidase is used as a second terminal oxidase in the mitochondria, electrons are transferred directly from reduced ubiquinol to oxygen forming water [2]. This is not coupled to ATP synthesis and is not inhibited by cyanide, this pathway is a single step process [1]. In rice the transcript levels of the alternative oxidase are increased by low temperature [1].

Number of members: 27

[1] Ito Y, Saisho D, Nakazono M, Tsutsumi N, Hirai A; Medline: 98086211 "Transcript levels of tandem-arranged alternative oxidase genes in rice are increased by low temperature." Gene 1997;203:121-129.

[2] Li Q, Ritzel RG, McLean LL, McIntosh L, Ko T, Bertrand H, Nargang FE; Medline: 96366413 "Cloning and analysis of the alternative oxidase gene of *Neurospora crassa*." Genetics 1996;142:129-140.

5

834. (APH)

Protein kinases signatures and profile

Cross-reference(s): PS00107; PROTEIN_KINASE_ATP, PS00108;

10

PROTEIN_KINASE_ST, PS00109; PROTEIN_KINASE_TYR, PS00111;
PROTEIN_KINASE_DOM

Eukaryotic protein kinases [1 to 5] are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of these regions have been selected to build signature patterns. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme [6]; two signature patterns were derived for that region: one specific for serine/threonine kinases and the other for tyrosine kinases. A profile was developed which is based on the alignment in [1] and covers the entire catalytic domain.

25

Consensus pattern: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K [K binds ATP]

30

Sequences known to belong to this class detected by the pattern the majority of known protein kinases but it fails to find a number of them, especially viral kinases which are quite divergent in this region and are completely missed by this pattern.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) [D is an active site residue]

Sequences known to belong to this class detected by the pattern. Most serine/ threonine specific protein kinases with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below.

Consensus pattern: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3) [D is an active site residue] tyrosine specific protein kinases with the exception of human ERBB3 and mouse blk. This pattern will also detect most bacterial aminoglycoside phosphotransferases [8,9] and herpesviruses ganciclovir kinases [10]; which are proteins structurally and evolutionary related to protein kinases. Sequences known to belong to this class detected by the profile ALL₁ except for three viral kinases. This profile also detects receptor guanylate cyclases (see <PDOC00430>) and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed before. It also detects Arabidopsis thaliana kinase-like protein TMKL1 which seems to have lost its catalytic activity.

Note if a protein analyzed includes the two protein kinase signatures, the probability of it being a protein kinase is close to 100%. Note eukaryotic-type protein kinases have also been found in prokaryotes such as *Myxococcus xanthus* [11] and *Yersinia pseudotuberculosis*. Note the patterns shown above has been updated since their publication in [7]. Note this documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

References

- [1] Hanks S.K., Hunter T., FASEB J. 9:576-596(1995).
- [2] Hunter T., Meth. Enzymol. 200:3-37(1991).
- [3] Hanks S.K., Quinn A.M., Meth. Enzymol. 200:38-62(1991).
- [4] Hanks S.K., Curr. Opin. Struct. Biol. 1:369-383(1991).

- [5] Hanks S.K., Quinn A.M., Hunter T., Science 241:42-52(1988).
- [6] Knighton D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor, S.S., Sowadski J.M., Science 253:407-414(1991).
- [7] Bairoch A., Claverie J.-M., Nature 331:22(1988).
- 5 [8] Benner S., Nature 329:21-21(1987).
- [9] Kirby R., J. Mol. Evol. 30:489-492(1992).
- [10] Littler E., Stuart A.D., Chee M.S., Nature 358:160-162(1992).
- [11] Munoz-Dorado J., Inouye S., Inouye M., Cell 67:995-1006(1991).

10

835. (Asp_Glu_race)

Aspartate and glutamate racemases signatures

Cross-reference(s) PS00923; ASP_GLU_RACEMASE_1 PS00924;

15

ASP_GLU_RACEMASE_2

Aspartate racemase (EC 5.1.1.13) and glutamate racemase (EC 5.1.1.3) are two evolutionary related bacterial enzymes that do not seem to require a cofactor for their activity [1].

20

Glutamate racemase, which interconverts L-glutamate into D-glutamate, is required for the biosynthesis of peptidoglycan and some peptide-based antibiotics such as gramicidin S. In addition to characterized aspartate and glutamate racemases, this family also includes a hypothetical protein from *Erwinia carotovora* and one from *Escherichia coli* (ygeA). Two conserved cysteines are present in the sequence of these enzymes. They are expected to play a role in catalytic activity by acting as bases in proton abstraction from the substrate.

25

Signature patterns were developed for both cysteines.

Consensus pattern: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK]

Consensus pattern: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM]

30

- [1] Gallo K.A., Knowles J.R., Biochemistry 32:3981-3990(1993).

836. (ATP-sulfurylase)

ATP-sulfurylase

This family consists of ATP-sulfurylase or sulfate adenylyltransferase EC:2.7.7.4 some of which are part of a bifunctional polypeptide chain associated with adenosyl phosphosulphate (APS) kinase APS_kinase. Both enzymes are required for PAPS (phosphoadenosine-phosphosulfate) synthesis from inorganic sulphate [2]. ATP sulfurylase catalyses the synthesis of adenosine-phosphosulfate APS from ATP and inorganic sulphate [1].

Number of members: 37

[1] Kurima K, Warman ML, Krishnan S, Domowicz M, Krueger RC Jr, Deyrup A, Schwartz NB; Medline: 98337975 "A member of a family of sulfate-activating enzymes causes murine brachymorphism" [published erratum appears in Proc Natl Acad Sci U S A 1998 Sep 29;95(20):12071] Proc Natl Acad Sci U S A 1998;95:8681-8685.

[2] Rosenthal E, Leustek T; Medline: 96096529 "A multifunctional Urechis caupo protein, PAPS synthetase, has both ATP sulfurylase and APS kinase activities." Gene 1995;165:243-248.

837. (ATP-synt_F)

ATP synthase (F/14-kDa) subunit

This family includes 14-kDa subunit from vATPases [1], which is in the peripheral catalytic part of the complex [2]. The family also includes archaeobacterial ATP synthase subunit F [3].

Number of members: 23

[1] Guo Y, Kaiser K, Wieczorek H, Dow JA; Medline: 96269411 "The Drosophila melanogaster gene vha14 encoding a 14-kDa F-subunit of the vacuolar ATPase." Gene 1996;172:239-243.

[2] Peng SB, Crider BP, Tsai SJ, Xie XS, Stone DK; Medline: 96216416 "Identification of a 14-kDa subunit associated with the catalytic sector of clathrin-coated vesicle H⁺-ATPase." J Biol Chem 1996;271:3324-3327.

[3] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968

- 5 "Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon *Methanosarcina mazei* Go1." J Biol Chem 1996;271:18843-18852.

838. (CBD_4)

10 Starch binding domain

Number of members: 48

15 839. (CbiX)

The function of CbiX is uncertain, however it is found in cobalamin biosynthesis operons and so may have a related function. Some CbiX proteins contain a striking histidine-rich region at their C-terminus, which suggests that it might be involved in metal chelation [1].

20 Number of members: 6

[1] Raux E, Lanois A, Warren MJ, Rambach A, Thermes C; Medline: 98416126 "Cobalamin (vitamin B12) biosynthesis: identification and characterization of a *Bacillus megaterium* cobI operon." Biochem J 1998;335:159-166.

840. (Complex1_51K)

30 Respiratory-chain NADH dehydrogenase 51 Kd subunit signatures Cross-reference(s)
PS00644; COMPLEX1_51K_1 PS00645; COMPLEX1_51K_2

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 51 Kd (in mammals), which is the second largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind to NAD, FMN, and a 2Fe-2S cluster.

The 51 Kd subunit is highly similar to [3,4]:

- Subunit alpha of *Alcaligenes eutrophus* NAD-reducing hydrogenase (gene *hoxF*) which also binds to NAD, FMN, and a 2Fe-2S cluster.
- Subunit NQO1 of *Paracoccus denitrificans* NADH-ubiquinone oxidoreductase.
- Subunit F of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoF*).

The 51 Kd subunit and the bacterial hydrogenase alpha subunit contains three regions of sequence similarities. The first one most probably corresponds to the NAD-binding site, the second to the FMN-binding site, and the third one, which contains three cysteines, to the iron-sulfur binding region. Signature patterns have been developed for the FMN-binding and for the 2Fe-2S binding regions.

Consensus pattern: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]- S
 Consensus pattern: E-S-C-G-x-C-x-P-C-R-x-G [The three C's are putative 2Fe-2S ligands]

[1] Ragan C.I., Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D., Eur. J. Biochem. 197:563-576(1991).

[3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).

[4] Weidner U., Geier S., Ptöck A., Friedrich T., Leif H., Weiss H., J. Mol. Biol. 233:109-122(1993).

841. (DAP_epimerase)

Diaminopimelate epimerase signature

Cross-reference(s) PS01326; DAP_EPIMERASE

Diaminopimelate epimerase (EC 5.1.1.7) catalyzes the isomerization of L,L- to D,L-meso-diaminopimelate in the biosynthetic pathway leading from aspartate to lysine. This enzyme is a protein of about 30 Kd. Two conserved cysteines seem [1] to function as the acid and base in the catalytic mechanism. As a signature pattern, the region surrounding the first of these two active site cysteines were selected.

Consensus pattern: N-x-D-G-S-x(4)-C-G-N-[GA]-x-R [C is an active site residue] Sequences known to belong to this class detected by the pattern ALL, except for an Anabaena dapF which has a Ser instead of the active site Cys.

[1] Cirilli M., Zheng R., Scapin G., Blanchard J.S., Biochemistry 37:16452-16458(1998).

842. (DNA_gyraseB_C)

DNA topoisomerase II signature

Cross-reference(s) PS00177; TOPOISOMERASE_II

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes *gyrA* and *gyrB* [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes *parC* and *parE*). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

<-----About-1400-residues----->

[-----Protein 39-*-----][----Protein 52----] Phage T4
[-----gyrB-----*-----][-----gyrA-----] Prokaryote II
 Archaeobacteria
[-----parE-----*-----][-----parD-----] Prokaryote IV
[-----*-----] Eukaryote and
 ASF

^{†*}: Position of the pattern.

As a signature pattern for this family of proteins, a region that contains a highly conserved pentapeptide was selected. The pattern is located in *gyrB*, in *parE*, and in protein 39 of phage T4 topoisomerase.

Consensus pattern: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG]

- [1] Sternglanz R., Curr. Opin. Cell Biol. 1:533-535(1990).
- [2] Bjornsti M.-A., Curr. Opin. Struct. Biol. 1:99-103(1991).
- [3] Sharma A., Mondragon A., Curr. Opin. Struct. Biol. 5:39-47(1995).
- [4] Roca J., Trends Biochem. Sci. 20:156-160(1995).

843. (DUF16)

Protein of unknown function

The function of this protein is unknown. It appears to only occur in *Mycoplasma pneumoniae*.

Number of members: 26

- [1] Himmelreich R, Hilbert H, Plagens H, Pirkel E, Li BC, Herrmann R; Medline: 97105885
 "Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*."
 Nucleic Acids Res 1996;24:4420-4449.

844. (DUF21)

5 Domain of unknown function

This transmembrane region has no known function. Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

15 845. (DUF56)

Integral membrane protein

The members of this family are putative integral membrane proteins. The function of the family is unknown, however the family includes Sec59 from yeast. Sec59 is a dolichol kinase EC:2.7.1.108, but it is not clear if the enzymatic activity resides in this region or its N terminal region.

Number of members: 13

846. (DUF94)

Domain of unknown function

The function of this domain is unknown. It is found in both eukaryotes and archaeobacteria. The alignment contains a completely conserved aspartate residue that may be functionally

694

important. The eukaryotic domains contains three conserved cysteines and a histidine that might be metal binding, however these are absent in the archaebacterial proteins.

Number of members: 9

5

847. (FF)

FF domain

10

This domain may be involved in protein-protein interaction [1].

Number of members: 42

15

[1] Bedford MT, Leder P; Medline: 99322199 "The FF domain: a novel motif that often accompanies WW domains." Trends Biochem Sci 1999;24:264-265.

848. (FLO_LFY)

20

Floricaula / Leafy protein

This family consists of various plant development proteins which are homologues of floricaula (FLO) and Leafy (LFY) proteins which are floral meristem identity proteins. Mutations in the sequences of these proteins affect flower and leaf development.

25

Number of members: 16

[1] Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N; Medline: 97411151 "UNIFOLIATA regulates leaf and flower morphogenesis in pea." Curr Biol 1997;7:581-587.

30

[2] Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM; Medline: 92274452 "LEAFY controls floral meristem identity in Arabidopsis." Cell 1992;69:843-859.

849. (G-patch)

G-patch domain

- 5 This domain is found in a number of RNA binding proteins, and is also found in proteins that contain RNA binding domains. This suggests that this domain may have an RNA binding function. This domain has seven highly conserved glycines.

Number of members: 47

10

[1] Aravind L, Koonin EV; Medline: 10470032 "G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins." Trends Biochem Sci 1999;24:342-344.

15

850. (Gram-ve_porins)

General diffusion Gram-negative porins signature

Cross-reference(s) PS00576; GRAM_NEG_PORIN

20

The outer membrane of Gram-negative bacteria acts as a molecular filter for hydrophilic compounds. Proteins, known as porins [1], are responsible for the 'molecular sieve' properties of the outer membrane. Porins form large water-filled channels which allows the diffusion of hydrophilic molecules into the periplasmic space. Some porins form general diffusion channels that allows any solutes up to a certain size (that size is known as the exclusion limit) to cross the membrane, while other porins are specific for a solute and contain a binding site for that solute inside the pores (these are known as selective porins). As porins are the major outer membrane proteins, they also serve as receptor sites for the binding of phages and bacteriocins. General diffusion porins generally assemble as trimer in the membrane and the transmembrane core of these proteins is composed exclusively of beta strands [2]. It has been shown [3] that a number of general porins are evolutionary related, these porins are:

30

- Enterobacteria phoE.
- Enterobacteria ompC.
- Enterobacteria ompF.

- Enterobacteria nmpC.
- Bacteriophage PA-2 LC.
- Neisseria PL.A.
- Neisseria PL.B.

5

As a signature pattern a conserved region was selected, located in the C-terminal part of these proteins, which spans two putative transmembrane beta strands.

Consensus pattern: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]- V

10

[1] Benz R., Bauer K., Eur. J. Biochem. 176:1-19(1988).

[2] Jap B.K., Walian P.J., Q. Rev. Biophys. 23:367-403(1990).

[3] Jeanteur D., Lakey J.H., Pattus F., Mol. Microbiol. 5:2153-2164(1991).

15

851. (HlyD)

HlyD family secretion proteins signature

Cross-reference(s) PS00543; HLYD_FAMILY

20

Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, require the help of two or more proteins for their secretion across the cell envelope. Amongst which a protein belonging to the ABC transporters family (see the relevant entry <PDOC00185>) and a protein belonging to a family which is currently composed [1 to 5] of the following members:

25

Gene	Species	Protein which is exported
------	---------	---------------------------

hlyD	Escherichia coli	Hemolysin
------	------------------	-----------

appD	A.pleuropneumoniae	Hemolysin
------	--------------------	-----------

30

lcnD	Lactococcus lactis	Lactococcin A
------	--------------------	---------------

lktD	A.actinomycetemcomitans	Leukotoxin
------	-------------------------	------------

	Pasteurella haemolytica	
--	-------------------------	--

rtxD	A.pleuropneumoniae	Toxin-III
------	--------------------	-----------

cyaD *Bordetella pertussis* Calmodulin-sensitive adenylate cyclase-hemolysin (cyclolysin)

cvaA *Escherichia coli* Colicin V

prtE *Erwinia chrysanthemi* Extracellular proteases B and C

5 aprE *Pseudomonas aeruginosa* Alkaline protease

emrA *Escherichia coli* Drugs and toxins

yjcR *Escherichia coli* Unknown

These proteins are evolutionary related and consist of from 390 to 480 amino acid residues.

They seem to be anchored in the inner membrane by a N-terminal transmembrane region.

10 Their exact role in the secretion process is not yet known. The C-terminal section of these proteins is the best conserved region; a signature pattern from that region was derived.

Consensus pattern: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-[LIVMFYW](2)-x-[LIVMFYW](3)

15 Sequences known to belong to this class detected by the pattern ALL, except for emrA and yjcR.

References:

[1] Gilson L., Mahanty H.K., Kolter R., EMBO J. 9:3875-3884(1990).

20 [2] Letoffe S., Delepelaire P., Wandersman C., EMBO J. 9:1375-1382(1990).

[3] Stoddard G.W., Petzel J.P., van Belkum M.J., Kok J., McKay L.L., Appl. Environ. Microbiol. 58:1952-1961(1992).

[4] Duong F., Lazdunski A., Cami B., Murgier M., Gene 121:47-54(1992).

[5] Lewis K., Trends Biochem. Sci. 19:119-123(1994).

25

852. (IBR)

In Between Ring fingers

30 The IBR (In Between Ring fingers) domain is found to occur between pairs of ring fingers (zf-C3HC4). The function of this domain is unknown. This domain has also been called the C6HC domain and DRIL (for double RING finger linked) domain [2].

Number of members: 25

[1] Morett E, Bork P; Medline: 10366851 "A novel transactivation domain in parkin." Trends Biochem Sci 1999;24:229-231.

[2] van der Reijden BA, Erpelinck-Verschueren CA, Lowenberg B, Jansen JH; Medline: 99349709 "TRIADs: a new class of proteins with a novel cysteine-rich signature." Protein Sci 1999;8:1557-1561.

853. (IPPT)

10 IPP transferase

[1] Durand JM, Bjork GR, Kuwae A, Yoshikawa M, Sasakawa C; Medline: 97440126 "The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of *Shigella flexneri* is required for expression of virulence genes." J Bacteriol 1997;179:5777-5782.

[2] Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC, Hopper AK; Medline: 94187700 "Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol." Mol Cell Biol 1994;14:2298-2306.

[3] Gillman EC, Slusher LB, Martin NC, Hopper AK; Medline: 91203856 "MOD5 translation initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA." Mol Cell Biol 1991;11:2382-2390.

854. (KE2)

25 KE2 family protein

The function of members of this family is unknown, although they have been suggested to contain a DNA binding leucine zipper motif [2].

30 Number of members: 9

[1] Ha H, Abe K, Artzt K; Medline: 92084131 "Primary structure of the embryo-expressed gene KE2 from the mouse H-2K region." Gene 1991;107:345-346.

[2] Shang HS, Wong SM, Tan HM, Wu M; Medline: 95129859 "YKE2, a yeast nuclear gene encoding a protein showing homology to mouse KE2 and containing a putative leucine-zipper motif." *Gene* 1994;151:197-201.

5

855. (Lipoprotein_6)

Prokaryotic membrane lipoprotein lipid attachment site

Cross-reference(s) PS00013; PROKAR_LIPOPROTEIN

10 In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- 15 - Major outer membrane lipoprotein (murein-lipoproteins) (gene *lpp*).
- *Escherichia coli* lipoprotein-28 (gene *nlpA*).
- *Escherichia coli* lipoprotein-34 (gene *nlpB*).
- *Escherichia coli* lipoprotein *nlpC*.
- *Escherichia coli* lipoprotein *nlpD*.
- 20 - *Escherichia coli* osmotically inducible lipoprotein B (gene *osmB*).
- *Escherichia coli* osmotically inducible lipoprotein E (gene *osmE*).
- *Escherichia coli* peptidoglycan-associated lipoprotein (gene *pal*).
- *Escherichia coli* rare lipoproteins A and B (genes *rplA* and *rplB*).
- *Escherichia coli* copper homeostasis protein *cutF* (or *nlpE*).
- 25 - *Escherichia coli* plasmids *traT* proteins.
- *Escherichia coli* Col plasmids lysis proteins.
- A number of *Bacillus* beta-lactamases.
- *Bacillus subtilis* periplasmic oligopeptide-binding protein (gene *oppA*).
- *Borrelia burgdorferi* outer surface proteins A and B (genes *ospA* and *ospB*).
- 30 - *Borrelia hermsii* variable major protein 21 (gene *vmp21*) and 7 (gene *vmp7*).
- *Chlamydia trachomatis* outer membrane protein 3 (gene *omp3*).
- *Fibrobacter succinogenes* endoglucanase *cel-3*.
- *Haemophilus influenzae* proteins *Pal* and *Pcp*.

- *Klebsiella pullulunase* (gene *pulA*).
- *Klebsiella pullulunase* secretion protein *pulS*.
- *Mycoplasma hyorhinis* protein *p37*.
- *Mycoplasma hyorhinis* variant surface antigens A, B, and C (genes *vlpABC*).
- 5 - *Neisseria* outer membrane protein *H.8*.
- *Pseudomonas aeruginosa* lipopeptide (gene *lppL*).
- *Pseudomonas solanacearum* endoglucanase *egl*.
- *Rhodopseudomonas viridis* reaction center cytochrome subunit (gene *cytC*).
- *Rickettsia* 17 Kd antigen.
- 10 - *Shigella flexneri* invasion plasmid proteins *mxj* and *mxIM*.
- *Streptococcus pneumoniae* oligopeptide transport protein A (gene *amiA*).
- *Treponema pallidum* 34 Kd antigen.
- *Treponema pallidum* membrane protein A (gene *tmpA*).
- *Vibrio harveyi* chitinase (gene *chb*).
- 15 - *Yersinia* virulence plasmid protein *ycsJ*.
- Halocyanin from *Natrobacterium pharaonis* [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion).

From the precursor sequences of all these proteins, a consensus pattern and a set of rules to identify this type of post-translational modification were derived.

Consensus pattern: {DERK}(6)-[LIVFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1)

25 The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

References

- [1] Hayashi S., Wu H.C., J. Bioenerg. Biomembr. 22:451-471(1990).
- [2] Klein P., Somorjai R.L., Lau P.C.K., Protein Eng. 2:15-20(1988).

[3] von Heijne G., Protein Eng. 2:531-534(1989).

[4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterheld D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

5

856. (Lipoprotein_7)

Adhesin lipoprotein

10

This family consists of the p50 and variable adherence-associated antigen (Vaa) adhesins from *Mycoplasma hominis*. *M. hominis* is a mycoplasma associated with human urogenital diseases, pneumonia, and septic arthritis [1]. An adhesin is a cell surface molecule that mediates adhesion to other cells or to the surrounding surface or substrate. The Vaa antigen is a 50-kDa surface lipoprotein that has four tandem repetitive DNA sequences encoding a periodic peptide structure, and is highly immunogenic in the human host [1]. p50 is also a 50-kDa lipoprotein, having three repeats A,B and C, that may be a tetramer of 191-kDa in its native environment [2].

15

Number of members: 18

20

[1] Zhang Q, Wise KS; Medline: 96294788 "Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent vaa genes. " Infect Immun 1996;64:2737-2744.

25

[2] Henrich B, Kitzerow A, Feldmann RC, Schaal H, Hadding U; Medline: 97047675 "Repetitive elements of the *Mycoplasma hominis* adhesin p50 can be differentiated by monoclonal antibodies." Infect Immun 1996;64:4027-4034.

30

857. (MaoC_like)

MaoC like domain

The MaoC protein is found to share similarity with a wide variety of enzymes; estradiol 17 beta-dehydrogenase 4, peroxisomal hydratase-dehydrogenase-epimerase, fatty acid synthase beta subunit. All these enzymes contain other domains. This domain is also present in the

NodN nodulation protein N. No specific function has been assigned to this region of any of these proteins. The maoC gene is part of a operon with maoA which is involved in the synthesis of monoamine oxidase [1].

5 Number of members: 46

[1] Sugino H, Sasaki M, Azakami H, Yamashita M, Murooka Y Medline: 96235221 "A monoamine-regulated *Klebsiella aerogenes* operon containing the monoamine oxidase structural gene (maoA) and the maoC gene." J Bacteriol 1992;174:2485-2492.

10

858. (MSP)

Manganese-stabilizing protein / photosystem II polypeptide

15 This family consists of the 33 KDa photosystem II polypeptide from the oxygen evolving complex (OEC) of plants and cyanobacteria. The protein is also known as the manganese-stabilizing protein as it is associated with the manganese complex of the OEC and may provide the ligands for the complex [1].

20 Number of members: 17

[1] Philbrick JB, Zilinskas BA; Medline: 88334494 "Cloning, nucleotide sequence and mutational analysis of the gene encoding the Photosystem II manganese-stabilizing polypeptide of *Synechocystis* 6803." Mol Gen Genet 1988;212:418-425.

25

859. (NAC)

[1] Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV; Medline: 99342100 "Comparative genomics of the Archaea (Euryarchacota): evolution of conserved protein families, the stable core, and the variable shell." Genome Res 1999;9:608-628.

30

Number of members: 27

860. (Nop)

5 Putative snoRNA binding domain

This family consists of various Pre RNA processing ribonucleoproteins. The function of the aligned region is unknown however it may be a common RNA or snoRNA or Nop1p binding domain. Nop5p (Nop58p) Swiss:Q12499 from yeast is the protein component of a ribonucleoprotein protein required for pre-18s rRNA processing and is suggested to function with Nop1p in a snoRNA complex [1]. Nop56p Swiss:O00567 and Nop5p interact with Nop1p and are required for ribosome biogenesis [2]. Prp31p Swiss:p49704 is required for pre-mRNA splicing in *S. cerevisiae* [3].

15 Number of members: 23

[1] Wu P, Brockenbrough JS, Metcalfe AC, Chen S, Aris JP; Medline: 98298165 "Nop5p is a small nucleolar ribonucleoprotein component required for pre- 18 S rRNA processing in yeast." *J Biol Chem* 1998;273:16453-16463.

20 [2] Gautier T, Berges T, Tollervy D, Hurt E; Medline: 8038777 "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis." *Mol Cell Biol* 1997;17:7088-7098.

25 [3] Weidenhammer EM, Singh M, Ruiz-Noriega M, Woolford JL Jr; Medline: 96184869 "The PRP31 gene encodes a novel protein required for pre-mRNA splicing in *Saccharomyces cerevisiae*." *Nucleic Acids Res* 1996;24:1164-1170.

861. (Nramp)

Natural resistance-associated macrophage protein

30

The natural resistance-associated macrophage protein (NRAMP) family consists of Nramp1, Nramp2, and yeast proteins Smf1 and Smf2. The NRAMP family is a novel family of functional related proteins defined by a conserved hydrophobic core of ten transmembrane

domains [5]. This family of membrane proteins are divalent cation transporters. Nramp1 is an integral membrane protein expressed exclusively in cells of the immune system and is recruited to the membrane of a phagosome upon phagocytosis [1]. By controlling divalent cation concentrations Nramp1 may regulate the interphagosomal replication of bacteria [1].

5 Mutations in Nramp1 may genetically predispose an individual to susceptibility to diseases including leprosy and tuberculosis conversely this might however provide protection from rheumatoid arthritis [1]. Nramp2 is a multiple divalent cation transporter for Fe²⁺, Mn²⁺ and Zn²⁺ amongst others it is expressed at high levels in the intestine; and is major transferrin-independent iron uptake system in mammals [1]. The yeast proteins Smf1 and Smf2 may also

10 transport divalent cations [3].

Number of members: 36

- [1] Govoni G, Gros P; Medline: 98383996 "Macrophage NRAMP1 and its role in resistance to microbial infections." *Inflamm Res* 1998;47:277-284.
- [2] Agranoff DD, Krishna S Medline: 98294035 "Metal ion homeostasis and intracellular parasitism." *Mol Microbiol* 1998;28:403-412.
- [3] Pinner E, Gruenheid S, Raymond M, Gros P; Medline: 98030569 "Functional complementation of the yeast divalent cation transporter family SMF by NRAMP2, a member of the mammalian natural resistance- associated macrophage protein family." *J Biol Chem* 1997;272:28933-28938.
- [4] Cellier M, Belouchi A, Gros P; Medline: 96402487 "Resistance to intracellular infections: comparative genomic analysis of Nramp." *Trends Genet* 1996;12:201-204.
- [5] Cellier M, Prive G, Belouchi A, Kwan T, Rodrigues V, Chia W, Gros P; Medline: 96036029 "Nramp defines a family of membrane proteins." *Proc Natl Acad Sci U S A* 1995;92:10089-10093.

862. (NTP_transf_2)

Nucleotidyltransferase domain

Members of this family belong to a large family of nucleotidyltransferases [1].

Number of members: 83

[1] Holm L, Sander C; Medline: 96005605 "DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily." Trends Biochem Sci 1995;20:345-347.

863. (Paramyxo_P)

Paramyxovirus P phosphoprotein

This family consists of paramyxovirus P phosphoprotein from sendai virus and human and bovine parainfluenza viruses. The P protein is an essential part of the viral RNA polymerase complex formed from the P and L proteins [1]. The exact role of the P protein in this complex is unknown but it is involved in multiple protein-protein interactions and binding the polymerase complex to the nucleocapsid or ribonucleoprotein template [1]. It also appears to be important for the proper folding of the L protein [1]. The paramyxoviruses have a negative sense ssRNA genome [1].

Number of members: 15

[1] Bowman MC, Smallwood S, Moyer SA; Medline: 99329169 "Dissection of Individual Functions of the Sendai Virus Phosphoprotein in Transcription." J Virol 1999;73:6474-6483.
[2] Matsuoka Y, Curran J, Pelet T, Kolakofsky D, Ray R, Compans RW; Medline: 91237868 "The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine-rich V protein." J Virol 1991;65:3406-3410.

864. (Patatin)

This family consists of various patatin glycoproteins from plants. The patatin protein accounts for up to 40% of the total soluble protein in potato tubers [2]. Patatin is a storage protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids [2].

Number of members: 21

[1] Banfalvi Z, Kostyal Z, Barta E; Medline: 95107249 "Solanum brevidens possesses a non-sucrose-inducible patatin gene." Mol Gen Genet 1994;245:517-522.

- 5 [2] Mignery GA, Pikaard CS, Park WD; Medline: 88226014 "Molecular characterization of the patatin multigene family of potato." Gene 1988;62:27-44.

865. (Pentapeptide_2)

- 10 Pentapeptide repeats (8 copies)

These repeats are found in many mycobacterial proteins. These repeats are most common in the PPE family of proteins, where they are found in the MPTR subfamily of PPE proteins.

- 15 The function of these repeats is unknown. The repeat can be approximately described as XNXGX, where X can be any amino acid. These repeats are similar to Pentapeptide [1], however it is not clear if these two families are structurally related.

Number of members: 362

[1] Bateman A, Murzin A, Teichmann SA; Medline: 98318059 "Structure and distribution of pentapeptide repeats in bacteria." Protein Sci 1998;7:1477-1480.

- 25 [2] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG; Medline: 98295987 "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." Nature 1998;393:537-544.

- 30 866. (Peptidase_C13)

Peptidase C13 family

This family of peptidases is known as the hemoglobinase family because it contains a globin degrading enzyme from blood parasites Swiss:P42665. However relatives are found in plants and other organisms that have other functions. Members of this family are asparaginyl peptidases [1].

5

Number of members: 26

10

[1] Chen JM, Dando PM, Rawlings ND, Brown MA, Young NE, Stevens RA, Hewitt E, Watts C, Barrett AJ; Medline: 97218252 "Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase." J Biol Chem 1997;272:8090-8098.

15

867. (Pro_dh)

Proline dehydrogenase

20

Number of members: 25

25

[1] Ling M, Allen SW, Wood JM; Medline: 95055736 "Sequence analysis identifies the proline dehydrogenase and delta 1- pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli PutA protein." J Mol Biol 1994;243:950-956.

30

868. (PsbP)

This family consists of the 23 kDa subunit of oxygen evolving system of photosystem II or PsbP from various plants (where it is encoded by the nuclear genome) and Cyanobacteria. The 23 KDa PsbP protein is required for PSII to be fully operational in vivo, it increases the affinity of the water oxidation site for Cl- and provides the conditions required for high affinity binding of Ca2+ [2].

Number of members: 25

[1] Rova EM, Mc Ewen B, Fredriksson PO, Styring S; Medline: 97067138 "Photoactivation and photoinhibition are competing in a mutant of *Chlamydomonas reinhardtii* lacking the 23-kDa extrinsic subunit of photosystem II." *J Biol Chem* 1996;271:28918-28924.

- 5 [2] Kochhar A, Khurana JP, Tyagi AK; Medline: 97191538 "Nucleotide sequence of the psbP gene encoding precursor of 23-kDa polypeptide of oxygen-evolving complex in *Arabidopsis thaliana* and its expression in the wild-type and a constitutively photomorphogenic mutant." *DNA Res* 1996;3:277-285.

10 869. (PUA)

The PUA domain named after PseudoUridine synthase and Archaeosine transglycosylase, was detected in archaeal and eukaryotic pseudouridine synthases, archaeal archaeosine synthases, a family of predicted ATPases that may be involved in RNA modification, a family of predicted archaeal and bacterial rRNA methylases. Additionally, the PUA domain was detected in a family of eukaryotic proteins that also contain a domain homologous to the translation initiation factor eIF1/SUI1; these proteins may comprise a novel type of translation factors. Unexpectedly, the PUA domain was detected also in bacterial and yeast glutamate kinases; this is compatible with the demonstrated role of these enzymes in the regulation of the expression of other genes [1]. It is predicted that the PUA domain is an RNA binding domain.

Number of members: 48

- 25 [1] Aravind L, Koonin EV; Medline: 99193178 "Novel predicted RNA-binding domains associated with the translation machinery." *J Mol Evol* 1999;48:291-302.

870. (RF1)

30 eRF1-like proteins

Members of this family are peptide chain release factors. The eukaryotic Release Factor 1 proteins (eRF1s) are involved in termination of translation. The eRF1 protein is functional for

all stop codons and appears to abolish read-through of these codons. This family also includes other proteins for which the precise molecular function is unknown. Many of them are from Archaeobacteria. These proteins may also be involved in translation termination but this awaits experimental verification. Number of members: 25

[1] Frolova L, Le Goff X, Rasmussen HH, Cheperegin S, Drugeon G, Kress M, Arman I, Haenni AL, Celis JE, Philippe M, et al; Medline: 95082951 "A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor" [see comments] Nature 1994;372:701-703.

[2] Drugeon G, Jean-Jean O, Frolova L, Le Goff X, Philippe M, Kisselev L, Haenni AL; Medline: 97315314 "Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA." Nucleic Acids Res 1997;25:2254-2258.

871. (Ribosomal_L14e)Ribosomal protein L14

This family includes the eukaryotic ribosomal protein L14.

Number of members: 15

872. (Ribosomal_S27)

Ribosomal protein S27a

This family of ribosomal proteins consists mainly of the 40S ribosomal protein S27a which is synthesized as a C-terminal extension of ubiquitin (CEP). The S27a domain comprises the C-terminal half of the protein. The synthesis of ribosomal proteins as extensions of ubiquitin promotes their incorporation into nascent ribosomes by a transient metabolic stabilization and is required for efficient ribosome biogenesis [3]. The ribosomal extension protein S27a contains a basic region that is proposed to form a zinc finger; its fusion gene is proposed as a mechanism to maintain a fixed ratio between ubiquitin necessary for degrading proteins and ribosomes a source of proteins [2].

Number of members: 36

873. (Spermine_synth)

5 Spermine/spermidine synthase

Spermine and spermidine are polyamines. This family includes spermidine synthase that catalyses the fifth (last) step in the biosynthesis of spermidine from arginine, and spermine synthase.

10

Number of members: 39

[1] Mezquita J, Pau M, Mezquita C; Medline: 97449308 "Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids." Gene 1997;195:313-319.

15

[2] Redman KL, Rechsteiner M; Medline: 89181932 "Identification of the long ubiquitin extension as ribosomal protein S27a." Nature 1989;338:438-440.

[3] Finley D, Bartel B, Varshavsky A; Medline: 89181925 "The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis." Nature 1989;338:394-401.

20

874. (Surp)

Surp module

25

[1] Denhez F, Lafyatis R; Medline: 94266805 "Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot." J Biol Chem 1994;269:16170-16179.

30

This domain is also known as the SWAP domain. SWAP stands for Suppressor-of-White-APricot. It has been suggested that these domains may be RNA binding [1].

Number of members: 32

875. (TFIIE)

TFIIE alpha subunit

The general transcription factor TFIIE has an essential role in eukaryotic transcription initiation together with RNA polymerase II and other general factors. Human TFIIE consists of two subunits TFIIE-alpha Swiss:P29083 and TFIIE-beta Swiss:P29084 and joins the preinitiation complex after RNA polymerase II and TFIIF [1]. This family consists of the conserved amino terminal region of eukaryotic TFIIE-alpha [2] and proteins from archaeobacteria that are presumed to be TFIIE-alpha subunits also Swiss:O29501 [3].

Number of members: 12

[1] Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder RG; Medline: 92065982 "Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE." Nature 1991;354:398-401.

[2] Ohkuma Y, Hashimoto S, Roeder RG, Horikoshi M; Medline: 93087200 Identification of two large subdomains in TFIIE-alpha on the basis of homology between Xenopus and human sequences. Nucleic Acids Res 1992;20:5838-5838.

[3] Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Venter JC, et al; Medline: 98049343 "The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon Archaeoglobus fulgidus." Nature 1997;390:364-370.

876. (Transglut_core)

Cross-reference(s) PS00547; TRANSGLUTAMINASES

Transglutaminases (EC 2.3.2.13) (TGase) [1,2] are calcium-dependent enzymes that catalyze the cross-linking of proteins by promoting the formation of isopeptide bonds between the gamma-carboxyl group of a glutamine in one polypeptide chain and the epsilon-amino group of a lysine in a second polypeptide chain. TGases also catalyze the conjugation of polyamines to proteins. The best known transglutaminase is blood coagulation factor XIII, a plasma tetrameric protein composed of two catalytic A subunits and two non-catalytic B subunits. Factor XIII is responsible for cross-linking fibrin chains, thus stabilizing the fibrin clot. Other forms of transglutaminases are widely distributed in various organs, tissues and body fluids. Sequence data is available for the following forms of TGase:

- Transglutaminase K (Tgase K), a membrane-bound enzyme found in mammalian epidermis and important for the formation of the cornified cell envelope (gene TGM1).
- Tissue transglutaminase (TGase C), a monomeric ubiquitous enzyme located in the cytoplasm (gene TGM2).
- Transglutaminase 3, responsible for the later stages of cell envelope formation in the epidermis and the hair follicle (gene TGM3).
- Transglutaminase 4 (gene TGM4).

A conserved cysteine is known to be involved in the catalytic mechanism of TGases. The erythrocyte membrane band 4.2 protein, which probably plays an important role in regulating the shape of erythrocytes and their mechanical properties, is evolutionary related to TGases. However the active site cysteine is substituted by an alanine and the 4.2 protein does not show TGase activity.

Consensus pattern:[GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G [The first C is the active site residue] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT/NONE.

[1] Ichinose A., Bottenus R.E., Davie E.W. J. Biol. Chem. 265:13411-13414(1990).

[2] Greenberg C.S., Birckbichler P.J., Rice R.H. FASEB J. 5:3071-3077(1991).

877. (TruB_N)

TruB family pseudouridylate synthase (N terminal domain)

Members of this family are involved in modifying bases in RNA molecules. They carry out the conversion of uracil bases to pseudouridine. This family includes TruB, a pseudouridylate synthase that specifically converts uracil 55 to pseudouridine in most tRNAs. This family also includes Cbf5p that modifies rRNA [2].

Number of members: 33

[1] Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J; Medline: 96079944 "Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli." RNA 1995;1:102-112.

[2] Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D; Medline: 98139521 "The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase." Genes Dev 1998;12:527-537.

878. (UDPGP)

UTP--glucose-1-phosphate uridylyltransferase

This family consists of UTP--glucose-1-phosphate uridylyltransferases, EC:2.7.7.9. Also known as UDP-glucose pyrophosphorylase (UDPGP) and Glucose-1-phosphate uridylyltransferase. UTP--glucose-1-phosphate uridylyltransferase catalyses the interconversion of MgUTP + glucose-1-phosphate and UDP-glucose + MgPPi [1]. UDP-glucose is an important intermediate in mammalian carbohydrate interconversion involved in various metabolic roles depending on tissue type [1]. In Dictyostelium (slime mold) mutants in this enzyme abort the development cycle [2]. Also within the family is UDP-N-acetylglucosamine Swiss:Q16222 or AGX1 [3] and two hypothetical proteins from Borrelia burgdorferi the lyme disease spirochaete Swiss:O51893 and Swiss:O51036.

Number of members: 18

[1] Duggleby RG, Chao YC, Huang JG, Peng HL, Chang HY; Medline: 96202932 "Sequence differences between human muscle and liver cDNAs for UDPglucose pyrophosphorylase and kinetic properties of the recombinant enzymes expressed in *Escherichia coli*." *Eur J Biochem* 1996;235:173-179.

5 [2] Ragheb JA, Dottin RP; Medline: 87231075 "Structure and sequence of a UDP glucose pyrophosphorylase gene of *Dictyostelium discoideum*." *Nucleic Acids Res* 1987;15:3891-3906.

[3] Mio T, Yabe T, Arisawa M, Yamada-Okabe H; Medline: 98269105 "The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression, and
10 catalytic mechanism. *J Biol Chem* 1998;273:14392-14397.

879. (UPF004)

Uncharacterized protein family UPF0044 signature

15 Cross-reference(s) PS01301; UPF0044

The following uncharacterized proteins have been shown [1] to be highly similar:

- *Bacillus subtilis* hypothetical protein yqel.
- *Escherichia coli* hypothetical protein yhbY and HU1333, the corresponding *Haemophilus influenzae* protein.
- *Methanococcus jannaschii* hypothetical protein MJ0652.

20 These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

25 Consensus pattern: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT/NONE.

30 880. (zf-A20)

A20-like zinc finger

A20- (an inhibitor of cell death)-like zinc fingers. The zinc

finger mediates self-association in A20. These fingers also mediate IL-1-induced NF-kappa B activation.

Number of members: 22

5

[1] Heyninck K, Beyaert R; Medline: 99126071 "The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF- kappaB activation at the level of TRAF6. FEBS Lett 1999;442:147-150.

10

[2] De Valck D, Heyninck K, Van Crielinge W, Contreras R, Beyaert R, Fiers W; Medline: 96390831 "A20, an inhibitor of cell death, self-associates by its zinc finger domain." FEBS Lett 1996;384:61-64.

6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

15

[3] Song HY, Rothe M, Goeddel DV; Medline: 96270609 "The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. Proc Natl Acad Sci U S A 1996;93:6721-6725.

[4] Opipari AW Jr, Boguski MS, Dixit VM; Medline: 90368626 "The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein." J Biol Chem 1990;265:14705-14708.

20

881. (zf-PARP)
Poly(ADP-ribose) polymerase zinc finger domain

Cross-reference(s) PS00347; PARP_ZN_FINGER_1 PS50064; PARP_ZN_FINGER_2

25

Poly(ADP-ribose) polymerase (EC 2.4.2.30) (PARP) [1,2] is a eukaryotic enzyme that catalyzes the covalent attachment of ADP-ribose units from NAD(+) to various nuclear acceptor proteins. This post-translational modification of nuclear proteins is dependent on DNA. It appears to be involved in the regulation of various important cellular processes such as differentiation, proliferation and tumor transformation as well as in the regulation of the molecular events involved in the recovery of the cell from DNA damage. Structurally, PARP, about 1000 amino-acids residues long, consists of three distinct domains: an N-terminal zinc-dependent DNA-binding domain, a central automodification domain and a C-terminal NAD-binding domain. The DNA-binding region contains a pair of

30

zinc finger domains which have been shown to bind DNA in a zinc-dependent manner. The zinc finger domains of PARP seem to bind specifically to single-stranded DNA. DNA ligase III [3] contains, in its N-terminal section, a single copy of a zinc finger highly similar to those of PARP.

Consensus pattern: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C [The three C's and the H are zinc ligands] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT NONE. Sequences known to belong to this class detected by the profile ALL. Other sequence(s) detected in SWISS-PROT NONE.

Note: This documentation entry is linked to both signature patterns and a profile. As the profile is much more sensitive than the patterns, you should use it if you have access to the necessary software tools to do so.

- [1] Althaus F.R., Richter C.R. Mol. Biol. Biochem. Biophys. 37:1-126(1987).
- [2] de Murcia G., Menissier de Murcia J. Trends Biochem. Sci. 19:172-176(1994).
- [3] Wei Y.-F., Robins P., Carter K., Caldecott K., Pappin D.J.C., Yu G.-L., Wang R.-P., Shell B.K., Nash R.A., Schar P., Barnes D.E., Haseltine W.A., Lindahl T. Mol. Cell. Biol. 15:3206-3216(1995).

882. Adenylylsulfate kinase (APS_kinase)

Enzyme that catalyses the phosphorylation of adenylylsulfate to 3'-phosphoadenylylsulfate. This domain contains an ATP binding P-loop motif. Number of members: 34

- [1] MacRae IJ, Rose AB, Segel IH; Medline: 99003196 "Adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*. site-directed mutagenesis at putative phosphoryl-accepting and ATP P-loop residues. J Biol Chem 1998;273:28583-28589.

883. DNA polymerase family B signature DNA_POLYMERASE_B (DNA_pol_B)

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a

primer for the de novo synthesis of a DNA chain. On the basis of sequence similarity, a number of DNA polymerases have been grouped [1 to 7] under the designation of DNA polymerase family B. These are:

- Higher eukaryotes polymerases alpha.
- 5 - Higher eukaryotes polymerases delta.
- Yeast polymerase I/alpha (gene POL1), polymerase II/epsilon (gene POL2), polymerase III/delta (gene POL3) and polymerase REV3.
- Escherichia coli polymerase II (gene dinA or polB).
- Archaeobacterial polymerases.
- 10 - Polymerases of viruses from the herpesviridae family.
- Polymerases from Adenoviruses.
- Polymerases from Baculoviruses.
- Polymerases from Chlorella viruses.
- Polymerases from Poxviruses.
- 15 - Bacteriophage T4 polymerase.
- Podoviridae bacteriophages Phi-29, M2 and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
- Polymerases encoded on mitochondrial linear DNA plasmids in various fungi and plants (Kluyveromyces lactis pGKL1 and pGKL2, Agaricus bitorquis pEM, Ascobolus immersus pAI2, Claviceps purpurea pCLK1, Neurospora Kalilo and Maranhar, maize S-1, etc).
- 20

Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a conserved tetrapeptide with two aspartate residues. Its function is not yet known. However, it has been suggested [3] that it

25 may be involved in binding a magnesium ion. This conserved region was selected as a signature for this family of DNA polymerases.

Consensus pattern [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC]

Sequences known to belong to this class detected by the patternALL, except for yeast polymerase II/epsilon, Agaricus bitorquis pEM and Sulfolobus solfataricus polymerase II.

30

[1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J. Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).

- [2] Bernad A., Zaballos A., Salas M., Blanco L. EMBO J. 6:4219-4225(1987).
 [3] Argos P. Nucleic Acids Res. 16:9909-9916(1988).
 [4] Wang T.S.-F., Wong S.W., Korn D. FASEB J. 3:14-21(1989).
 [5] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Eng. 3:461-467(1990).
 [6] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).
 [7] Braithwaite D.K., Ito J. Nucleic Acids Res. 21:787-802(1993).

884. DNA polymerase family X signature - DNA_POLYMERASE_X (DNA_polymeraseX)

DNA polymerases (EC 2.7.7.7) can be classified, on the basis of sequence similarity [1], into at least four different groups: A, B, C and X. DNA polymerases that belong to family X are listed below [2]:

- Vertebrate polymerase beta, involved in DNA repair.
- Yeast polymerase IV (POL4) [3], an enzyme with similar characteristics to that of the mammalian polymerase beta.
- Terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31). TdT catalyzes the elongation of polydeoxynucleotide chains by terminal addition. One of the functions of this enzyme is the addition of nucleotides at the junction of rearranged Ig heavy chain and T cell receptor gene segments during the maturation of B and T cells.
- African Swine Fever virus protein O174L [4].
- Fission yeast hypothetical protein SpAC2F7.06c.

These enzymes are small (about 40 Kd) compared with other polymerases and their reaction mechanism operates via a distributive mode, i.e. they dissociate from the template-primer after addition of each nucleotide.

As a signature pattern for this family of DNA polymerases, a highly conserved region that contains a conserved arginine and two conserved aspartic acid residues were selected. The latter together with the arginine have been shown [5] to be involved in primer binding in polymerase beta.

Consensus pattern G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D- [LIVMFY](3)-x(2)-[SAP] Sequences known to belong to this class detected by the patternALL.

[1] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).

5 [2] Matsukage A., Nishikawa K., Ooi T., Seto Y., Yamaguchi M. J. Biol. Chem. 262:8960-8962(1987).

[3] Prasad R., Widen S.G., Singhal R.K., Watkins J., Prakash L., Wilson S.H. Nucleic Acids Res. 21:5301-5307(1993).

10 [4] Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinuela E. Virology 208:249-278(1995).

[5] Date T., Yamamoto S., Tanihara K., Nishimoto Y., Matsukage A. Biochemistry 30:5286-5292(1991).

885. DUF14 - Domain of unknown function

15 This domain is found in glutamate synthase, tungsten formylmethanofuran dehydrogenase subunit c (FwdC) and molybdenum formylmethanofuran dehydrogenase subunit c (FmdC). It has no known function. Number of members: 52

20 [1] Hochheimer A, Hedderich R, Thauer RK; Medline: 99035764. "The formylmethanofuran dehydrogenase isoenzymes in Methanobacterium wolfei and Methanobacterium thermoautotrophicum: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme." Arch Microbiol 1998;170:389-393.

886. DUF18-Domain of unknown function

25 This domain of unknown function is found in several C. elegans proteins. The domain is 120 amino acids long and rich in cysteine residues. There are 16 conserved cysteine positions in the domain. Number of members: 34

887. DUF27-Domain of unknown function

30 This domain is found in a number of otherwise unrelated proteins. This domain is found at the C-terminus of the macro-H2A histone protein Swiss:Q02874. This domain is found in the non-structural proteins of several types of ssRNA viruses such as NSP2 from alphaviruses Swiss:P03317. This domain is also found on its own in a family of proteins from bacteria

Swiss:P75918, archaeobacteria Swiss:O59182 and eukaryotes Swiss:Q17432, suggesting that it is involved in an important and ubiquitous cellular process. Number of members: 66

888. DUF37-Domain of unknown function

- 5 This domain is found in short (70 amino acid) hypothetical proteins from various bacteria. The domain contains three conserved cysteine residues. Swiss:Q44066 from *Aeromonas hydrophila* has been found to have hemolytic activity (unpublished). Number of members: 19

10 889. EGF-like domain signatures. (EGF-like)

A sequence of about thirty to forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown [1 to 6] to be present, in a more or less conserved form, in a large number of other, mostly animal proteins. The proteins currently known to contain one or more copies of an EGF-like pattern are listed below.

- 15 - Adipocyte differentiation inhibitor (gene PREF-1) from mouse (6 copies).
 - Agrin, a basal lamina protein that causes the aggregation of acetylcholine receptors on cultured muscle fibers (4 copies).
 - Amphiregulin, a growth factor (1 copy).
 - Betacellulin, a growth factor (1 copy).
 20 - Blastula proteins BP10 and Span from sea urchin which are thought to be involved in pattern formation (1 copy).
 - BM86, a glycoprotein antigen of cattle tick (7 copies).
 - Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity (1-2 copies). Homologous
 25 proteins are found in sea urchin - suBMP (1 copy) - and in *Drosophila* - the dorsal-ventral patterning protein tolloid (2 copies).
 - *Caenorhabditis elegans* developmental proteins lin-12 (13 copies) and glp-1 (10 copies).
 - *Caenorhabditis elegans* APX-1 protein, a patterning protein (4.5 copies).
 - Calcium-dependent serine proteinase (CASP) which degrades the extracellular matrix
 30 proteins type I and IV collagen and fibronectin (1 copy).
 - Cartilage matrix protein CMP (1 copy).
 - Cartilage oligomeric matrix protein COMP (4 copies).
 - Cell surface antigen 114/A10 (3 copies).

- Cell surface glycoprotein complex transmembrane subunit ASGP-2 from rat (2 copies).
- Coagulation associated proteins C, Z (2 copies) and S (4 copies).
- Coagulation factors VII, IX, X and XII (2 copies).
- Complement C1r components (1 copy).
- 5 - Complement C1s components (1 copy).
- Complement-activating component of Ra-reactive factor (RARF) (1 copy).
- Complement components C6, C7, C8 alpha and beta chains, and C9 (1 copy).
- Crumbs, an epithelial development protein from *Drosophila* (29 copies).
- Epidermal growth factor precursor (7-9 copies).
- 10 - Exogastrula-inducing peptides A, C, D and X from sea urchin (1 copy).
- Fat protein, a *Drosophila* cadherin-related tumor suppressor (5 copies).
- Fetal antigen 1, a probable neuroendocrine differentiation protein, which is derived from the delta-like protein (DLK) (6 copies).
- Fibrillin 1 (47 copies) and fibrillin 2 (14 copies).
- 15 - Fibropellins IA (21 copies), IB (13 copies), IC (8 copies), II (4 copies) and III (8 copies) from the apical lamina - a component of the extracellular matrix - of sea urchin.
- Fibulin-1 and -2, two extracellular matrix proteins (9-11 copies).
- Giant-lens protein (protein Argos), which regulates cell determination and axon guidance in the *Drosophila* eye (1 copy).
- 20 - Growth factor-related proteins from various poxviruses (1 copy).
- Gurken protein, a *Drosophila* developmental protein (1 copy).
- Heparin-binding EGF-like growth factor (HB-EGF), transforming growth factor alpha (TGF-alpha), growth factors Lin-3 and Spitz (1 copy); the precursors are membrane proteins, the mature form is located extracellular.
- 25 - Hepatocyte growth factor (HGF) activator (EC 3.4.21.-) (2 copies).
- LDL and VLDL receptors, which bind and transport low-density lipoproteins and very low-density lipoproteins (3 copies).
- LDL receptor-related protein (LRP), which may act as a receptor for endocytosis of extracellular ligands (22 copies).
- 30 - Leucocyte antigen CD97 (3 copies), cell surface glycoprotein EMR1 (6 copies) and cell surface glycoprotein F4/80 (7 copies).
- Limulus clotting factor C, which is involved in hemostasis and host defense mechanisms in japanese horseshoe crab (1 copy).

- Meprin A alpha subunit, a mammalian membrane-bound endopeptidase (1 copy).
- Milk fat globule-EGF factor 8 (MFG-E8) from mouse (2 copies).
- Neuregulin GGF-I and GGF-II, two human glial growth factors (1 copy).
- Neurexins from mammals (3 copies).
- 5 - Neurogenic proteins Notch, Xotch and the human homolog Tan-1 (36 copies), Delta (9 copies) and the similar differentiation proteins Lag-2 from *Caenorhabditis elegans* (2 copies), Serrate (14 copies) and Slit (7 copies) from *Drosophila*.
- Nidogen (also called entactin), a basement membrane protein from chordates (2-6 copies).
- Ookinete surface proteins (24 Kd, 25 Kd, 28 Kd) from *Plasmodium* (4 copies).
- 10 - Pancreatic secretory granule membrane major glycoprotein GP2 (1 copy).
- Perforin, which lyses non-specifically a variety of target cells (1 copy).
- Proteoglycans aggrecan (1 copy), versican (2 copies), perlecan (at least 2 copies), brevican (1 copy) and chondroitin sulfate proteoglycan (gene PG-M) (2 copies).
- Prostaglandin G/H synthase 1 and 2 (EC 1.14.99.1) (1 copy), which is found in the
- 15 endoplasmatic reticulum.
- S1-5, a human extracellular protein whose ultimate activity is probably modulated by the environment (5 copies).
- Schwannoma-derived growth factor (SDGF), an autocrine growth factor as well as a mitogen for different target cells (1 copy).
- 20 - Selectins. Cell adhesion proteins such as ELAM-1 (E-selectin), GMP-140 (P-selectin), or the lymph-node homing receptor (L-selectin) (1 copy).
- Serine/threonine-protein kinase homolog (gene Pro25) from *Arabidopsis thaliana*, which may be involved in assembly or regulation of light-harvesting chlorophyll A/B protein (2 copies).
- 25 - Sperm-egg fusion proteins PH-30 alpha and beta from guinea pig (1 copy).
- Stromal cell derived protein-1 (SCP-1) from mouse (6 copies).
- TDGF-1, human teratocarcinoma-derived growth factor 1 (1 copy).
- Tenascin (or neuronectin), an extracellular matrix protein from mammals (14.5 copies), chicken (TEN-A) (13.5 copies) and the related proteins human tenascin-X (18 copies) and
- 30 tenascin-like proteins TEN-A and TEN-M from *Drosophila* (8 copies).
- Thrombomodulin (fetomodulin), which together with thrombin activates protein C (6 copies).

- Thrombospondin 1, 2 (3 copies), 3 and 4 (4 copies), adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions.

- Thyroid peroxidase 1 and 2 (EC 1.11.1.8) from human (1 copy).

- Transforming growth factor beta-1 binding protein (TGF-B1-BP) (16 or 18 copies).

5 - Tyrosine-protein kinase receptors Tek and Tie (EC 2.7.1.112) (3 copies).

- Urokinase-type plasminogen activator (EC 3.4.21.73) (UPA) and tissue plasminogen activator (EC 3.4.21.68) (TPA) (1 copy).

- Uromodulin (Tamm-horsfall urinary glycoprotein) (THP) (3 copies).

10 - Vitamin K-dependent anticoagulants protein C (2 copies) and protein S (4 copies) and the similar protein Z, a single-chain plasma glycoprotein of unknown function (2 copies).

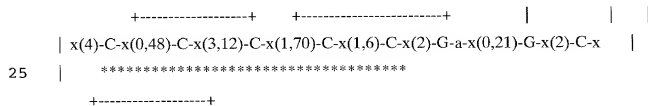
- 63 Kd sperm flagellar membrane protein from sea urchin (3 copies).

- 93 Kd protein (gene nel) from chicken (5 copies).

- Hypothetical 337.6 Kd protein T20G5.3 from *Caenorhabditis elegans* (44 copies).

15 The functional significance of EGF domains in what appear to be unrelated proteins is not yet clear. However, a common feature is that these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in disulfide bonds. The main structure is a two-stranded beta-sheet

20 followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines strongly vary in length as shown in the following schematic representation of the EGF-like domain:



'C': conserved cysteine involved in a disulfide bond.

'G': often conserved glycine

30 'a': often conserved aromatic amino acid

'*': position of both patterns.

'x': any residue

The region between the 5th and 6th cysteine contains two conserved glycines of which at least one is present in most EGF-like domains. Two patterns were created for this domain, each including one of these C-terminal conserved glycine residues.

- 5 Consensus pattern: C-x-C-x(5)-G-x(2)-C [The 3 C's are involved in disulfide bonds]
Sequences known to belong to this class detected by the pattern A majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT87 proteins, of which 27 can be considered as possible candidates.

10

Consensus pattern: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C [The three C's are involved in disulfide bonds] Sequences known to belong to this class detected by the pattern A majority, but not those that have very long or very short regions between the last 3 conserved cysteines of their EGF-like domain(s). Other sequence(s) detected in SWISS-PROT83 proteins, of which 49 can be considered as possible candidates. Note The beta chain of the integrin family of proteins contains 2 cysteine- rich repeats which were said to be dissimilar with the EGF pattern [7].

15

20

Note Laminin EGF-like repeats (see <PDOC00961>) are longer than the average EGF module and contain a further disulfide bond C-terminal of the EGF-like region. Perlecan and agrin contain both EGF-like domains and laminin-type EGF-like domains. Note the pattern do not detect all of the repeats of proteins with multiple EGF-like repeats. Note see <PDOC00913> for an entry describing specifically the subset of EGF- like domains that bind calcium.

25

[1] Davis C.G. New Biol. 2:410-419(1990).

[2] Blomquist M.C., Hunt L.T., Barker W.C. Proc. Natl. Acad. Sci. U.S.A. 81:7363-7367(1984).

[3] Barker W.C., Johnson G.C., Hunt L.T., George D.G. Protein Nucl. Acid Enz. 29:54-68(1986).

30

[4] Doolittle R.F., Feng D.F., Johnson M.S. Nature 307:558-560(1984).

[5] Appella E., Weber I.T., Blasi F. FEBS Lett. 231:1-4(1988).

[6] Campbell I.D., Bork P. Curr. Opin. Struct. Biol. 3:385-392(1993).

[7] Tamkun J.W., DeSimone D.W., Fonda D., Patel R.S., Buck C., Horwitz A.F., Hynes R.O. Cell 46:271-282(1986).

5 890. Ham1 family (Ham1p_like)

This family consists of the HAM1 protein Swiss:P47119 and hypothetical archaeal bacterial and *C. elegans* proteins. HAM1 controls 6-N-hydroxylaminopurine (HAP) sensitivity and mutagenesis in *S. cerevisiae* Swiss:P47119 [1]. The HAM1 protein protects the cell from HAP, either on the level of deoxynucleoside triphosphate or the DNA level by a yet
10 unidentified set of reactions [1]. Number of members: 19

[1] Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H, Pavlov YI; Medline: 96381244 "HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast *Saccharomyces cerevisiae*." Yeast 1996;12:17-29.

15 891. (HCO3_cotransp)

Anion exchange is a cellular transport function which contributes to the regulation of cell pH and volume. Anion exchangers are a family of functionally related proteins that contributes to
20 these properties by maintaining the intracellular level of the two principal anions: chloride and HCO3-. The best characterized anion exchanger is the band 3 protein [1], which is an erythrocyte anion exchange membrane glycoprotein. Band 3 is a protein of about 900 amino acids which consists of a cytoplasmic N-terminal domain of about 400 residues and an hydrophobic C-terminal section of about 500 residues that contains at least ten
25 transmembrane regions. The cytoplasmic domain provides binding sites for cytoskeletal proteins, while the integral membrane domain is responsible for anion transport. Band 3 protein is specific to erythroid cells, at least two other proteins [2] structurally and functionally related to band 3, are found in nonerythroid tissues:

- AE2 (or B3 related protein; B3RP), a protein of 1200 residues, which seems to be present
30 in a variety of cell types including lymphoid, kidney, and choroid plexus.

- AE3, a protein of 1200 residues, which is specific to neurons.

Structurally AE2 and AE3 are very similar to band 3, the main difference being an extension of some 300 residues of the N-terminal domain in AE2 and AE3.

Two signature patterns were developed for these proteins. The first pattern is based on a conserved stretch of sequence that contains four clustered positive charged residues and which is located at the C-terminal extremity of the cytoplasmic domain, just before the first transmembrane segment from the integral domain. The second pattern is based on the perfectly conserved sequence of the fifth transmembrane segment; this segment contains a lysine, which is the covalent binding site for the isothiocyanate group of DIDS, an inhibitor of anion exchange.

Consensus pattern F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L Sequences known to belong to this class detected by the pattern ALL.

- [1] Jay D., Cantley L. Annu. Rev. Biochem. 55:511-538(1986).
- [2] Reithmeier R.A.F. Curr. Opin. Struct. Biol. 3:515-523(1993).

892. ATP phosphoribosyltransferase signature (HisG)

ATP phosphoribosyltransferase (EC 2.4.2.17) is the enzyme that catalyzes the first step in the biosynthesis of histidine in bacteria, fungi and plants. It is a protein of about 23 to 32 Kd. As a signature pattern a region located in the C-terminal part of this enzyme was selected.

Consensus pattern E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM]
Sequences known to belong to this class detected by the pattern ALL.

893. HNH endonuclease (HNH)

Number of members: 56

- [1] Shub DA, Goodrich-Blair H, Eddy SR; Medline: 95117127 "Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns." Trends Biochem Sci 1994;19:402-404.

[2] Dalggaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854
 “Statistical modeling and analysis of the LAGLIDADG family of site- specific endonucleases
 and identification of an intein that encodes a site-specific endonuclease of the HNH family.”
 Nucleic Acids Res 1997;25:4626-4638.

[3] Gorbalenya AE; Medline: 95004046 “Self-splicing group I and group II introns encode
 homologous (putative) DNA endonucleases of a new family.” Protein Sci 1994;3:1117-1120.

894. NEUROHYPOPHYS_HORM (hormone5)

Oxytocin (or ocytocin) and vasopressin [1] are small (nine amino acid residues), structurally
 and functionally related neurohypophysial peptide hormones. Oxytocin causes contraction of
 the smooth muscle of the uterus and of the mammary gland while vasopressin has a direct
 antidiuretic action on the kidney and also causes vasoconstriction of the peripheral vessels.
 Like the majority of active peptides, both hormones are synthesized as larger protein
 precursors that are enzymatically converted to their mature forms. Peptides belonging to this
 family are also found in birds, fish, reptiles and amphibians (mesotocin, isotocin, valitocin,
 glutitocin, aspartocin, vasotocin, seritocin, asvatocin, phasvatocin), in worms (annetocin),
 octopi (cephalotocin), locust (locupressin or neuropeptide F1/F2) and in molluscs
 (conopressins G and S) [2]. The pattern developed to detect this category of peptides spans
 their entire sequence and includes four invariant amino acid residues.

Consensus pattern C-[LIFY](2)-x-N-[CS]-P-x-G [The two C's are linked by a disulfide
 bond]. Sequences known to belong to this class detected by the pattern ALL.

[1] Acher R., Chauvet J. Biochimie 70:1197-1207(1988).

[2] Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R. Int. J. Pept. Protein
 Res. 45:482-487(1995).

895. 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (HPPK)

All organisms require reduced folate cofactors for the synthesis of a variety of metabolites.
 Most microorganisms must synthesize folate de novo because they lack the active transport
 system of higher vertebrate cells which allows these organisms to use dietary folates.
 Enzymes involved in folate biosynthesis are therefore targets for a variety of antimicrobial

- 10 Consensus pattern [KRHD]-x-[GA]-[PSAE]-R-x(2)-D-[LIV]-D-[LIVM](2) Sequences
known to belong to this class detected by the pattern ALL. Other sequence(s) detected in
SWISS-PROT NONE.

- 20 896. Metalloenzyme superfamily (Metalloenzyme)
This family includes phosphopentomutase Swiss:P07651 and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, Swiss:P37689. This family is also related to alk_phosphatase [1]. The alignment contains the most conserved residues that are probably involved in metal binding and catalysis. Number of members: 34

- 897. Penicillin amidase (Penicil_amidase)**
Penicillin amidase or penicillin acylase EC:3.5.1.11 catalyses the hydrolysis of benzylpenicillin to phenylacetic acid and 6-aminopenicillanic acid (6-APA) a key

intermediate in the synthesis of penicillins [1]. Also in the family is cephalosporin acylase Swiss:P07662 and Swiss:P29958 aculeacin A acylase which are involved in the synthesis of related peptide antibiotics. Number of members: 13

- 5 [1] Verhaert RM, Riemens AM, van der Laan JM, van Duin J, Quax WJ; Medline: 97438505 "Molecular cloning and analysis of the gene encoding the thermostable penicillin G acylase from *Alcaligenes faecalis*. Appl Environ Microbiol 1997;63:3412-3418.
- [2] Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; Medline: 95115804 "Penicillin acylase has a single-amino-acid catalytic centre." Nature 1995;373:264-268.

10

898. Phosphoribosyl-AMP cyclohydrolase (PRA-CH)

This enzyme catalyses the third step in the histidine biosynthetic pathway. It requires Zn ions for activity. Number of members: 13

15

- [1] D'Ordine RL, Klem TJ, Davisson VJ; Medline: 99129952 "N1-(5'-phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: purification and characterization of a unique metalloenzyme. Biochemistry 1999;38:1537-1546.

20

899. Phosphoribosyl-ATP pyrophosphohydrolase (PRA-PH)

This enzyme catalyses the second step in the histidine biosynthetic pathway. Number of members: 32

- 25 [1] Keesey JK Jr, Bigelis R, Fink GR; Medline: 79216449 "The product of the *his4* gene cluster in *Saccharomyces cerevisiae*. A trifunctional polypeptide." J Biol Chem 1979 Aug 10;254:7427-7433.
- [2] Bruni CB, Carlomagno MS, Formisano S, Paoletta G; Medline: 86310274 "Primary and secondary structural homologies between the *HIS4* gene product of *Saccharomyces cerevisiae* and the *hisIE* and *hisD* gene products of *Escherichia coli* and *Salmonella typhimurium*." Mol Gen Genet 1986;203:389-396.

30

900. Prokaryotic membrane lipoprotein lipid attachment site (PstS)

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [1]. Some of the proteins known to undergo such processing currently include (for recent listings see [1,2,3]):

- Major outer membrane lipoprotein (murein-lipoproteins) (gene *lpp*).
- *Escherichia coli* lipoprotein-28 (gene *nlpA*).
- *Escherichia coli* lipoprotein-34 (gene *nlpB*).
- *Escherichia coli* lipoprotein *nlpC*.
- *Escherichia coli* lipoprotein *nlpD*.
- *Escherichia coli* osmotically inducible lipoprotein B (gene *osmB*).
- *Escherichia coli* osmotically inducible lipoprotein E (gene *osmE*).
- *Escherichia coli* peptidoglycan-associated lipoprotein (gene *pal*).
- *Escherichia coli* rare lipoproteins A and B (genes *rplA* and *rplB*).
- *Escherichia coli* copper homeostasis protein *cutF* (or *nlpE*).
- *Escherichia coli* plasmids *traT* proteins.
- *Escherichia coli* Col plasmids lysis proteins.
- A number of *Bacillus* beta-lactamases.
- *Bacillus subtilis* periplasmic oligopeptide-binding protein (gene *oppA*).
- *Borrelia burgdorferi* outer surface proteins A and B (genes *ospA* and *ospB*).
- *Borrelia hermsii* variable major protein 21 (gene *vmp21*) and 7 (gene *vmp7*).
- *Chlamydia trachomatis* outer membrane protein 3 (gene *omp3*).
- *Fibrobacter succinogenes* endoglucanase *cel-3*.
- *Haemophilus influenzae* proteins *Pal* and *Pcp*.
- *Klebsiella* pullulanase (gene *pulA*).
- *Klebsiella* pullulanase secretion protein *pulS*.
- *Mycoplasma hyorhinis* protein *p37*.
- *Mycoplasma hyorhinis* variant surface antigens A, B, and C (genes *vlpABC*).
- *Neisseria* outer membrane protein H.8.
- *Pseudomonas aeruginosa* lipopeptide (gene *lppL*).
- *Pseudomonas solanacearum* endoglucanase *egl*.
- *Rhodopseudomonas viridis* reaction center cytochrome subunit (gene *cytC*).

- Rickettsia 17 Kd antigen.
- Shigella flexneri invasion plasmid proteins mxiJ and mxiM.
- Streptococcus pneumoniae oligopeptide transport protein A (gene amiA).
- Treponema pallidum 34 Kd antigen.
- 5 - Treponema pallidum membrane protein A (gene tmpA).
- Vibrio harveyi chitobiase (gene chb).
- Yersinia virulence plasmid protein yscJ.
- Halocyanin from Natrobacterium pharaonis [4], a membrane associated copper-binding protein. This is the first archaeobacterial protein known to be modified in such a fashion).
- 10 From the precursor sequences of all these proteins, a consensus pattern was derived and a set of rules to identify this type of post-translational modification.

Consensus pattern {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C [C is the lipid attachment site] Additional rules: 1) The cysteine must be between positions 15 and 35 of the sequence in consideration. 2) There must be at least one Lys or one Arg in the first seven positions of the sequence. Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT some 100 prokaryotic proteins. Some of them are not membrane lipoproteins, but at least half of them could be.

- [1] Hayashi S., Wu H.C. J. Bioenerg. Biomembr. 22:451-471(1990).
- [2] Klein P., Somorjai R.L., Lau P.C.K. Protein Eng. 2:15-20(1988).
- [3] von Heijne G. Protein Eng. 2:531-534(1989).
- [4] Mattar S., Scharf B., Kent S.B.H., Rodewald K., Oesterhelt D., Engelhard M. J. Biol. Chem. 269:14939-14945(1994).

901. Ribosome recycling factor (RRF)

The ribosome recycling factor (RRF / ribosome release factor) dissociates the ribosome from the mRNA after termination of translation, and is essential bacterial growth [1]. Thus ribosomes are "recycled" and ready for another round of protein synthesis. Number of members: 27

[1] Janosi L, Shimizu I, Kaji A; Medline: 94240115 "Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth." Proc Natl Acad Sci U S A 1994;91:4249-4253.

5

902. S-layer homology (SLH)

S-layers are paracrystalline mono-layered assemblies of (glyco)proteins which coat the surface of bacteria [1]. Several S-layer proteins and some other cell wall proteins contain one or more copies of a domain of about 50-60 residues, which has been called SLH (for S-layer homology) [2]. There is strong evidence that this domain serves as an anchor to the peptidoglycan [3]. The SLH domain has been found in:

10

- S-layer glycoprotein of *Acetogenium kivui* (3 copies).

- S-layer 125 Kd protein of *Bacillus sphaericus* (3 copies).

- S-layer protein of *Bacillus anthracis* (3 copies).

15

- S-layer protein of *Bacillus licheniformis* (3 copies).

- S-layer protein (HWP) from *Bacillus brevis* strain HPD31 (3 copies).

- Middle cell wall protein (MWP) from *Bacillus brevis* strain 47 (3 copies).

- S-layer protein (p100) of *Thermus thermophilus* (1 copy).

20

- Outer membrane protein Omp-alpha from *Thermotoga maritima* (1 copy).

- Cellulosome anchoring protein (gene *ancA*), outer layer protein B (OlpB) and a further potential cell surface glycoprotein from *Clostridium thermocellum* (3 copies; the first copy is missing its N-terminal third which is appended to the end of the third copy; may have arisen by circular permutation).

- Amylopullulanase (gene *amyB*) from *Thermoanaerobacter thermosulfurogenes* (3 copies)

25

- Amylopullulanase (gene *aapT*) from *Bacillus* strain XAL-601 (3 copies).

- Endoglucanase from *Bacillus* strain KSM-635 (3 copies).

- Exoglucanase (gene *xynX*) from *Clostridium thermocellum* (3 copies).

- Xylanase A (gene *xynA*) from *Thermoanaerobacter saccharolyticum* (2 copies; 3 copies if a frameshift is taken into account).

30

- Protein involved in butirosin production (ButB) from *Bacillus circulans* (2 incomplete copies; 3 copies if three frameshifts are taken into account).

- Two hypothetical proteins from *Synechocystis* strain PCC 6803 (1 copy each).

- A hypothetical protein with sequence similarity to amylopullulanases found 3' of amylase gene from *Bacillus circulans* (fragment of 1 copy; 3 copies if two frameshifts are taken into account).

SLH domains are found at the N- or C-termini of mature proteins. They occur in single copy followed by a predicted coiled coil domain, or in three contiguous copies. Structurally, the SLH domain is predicted to contain two alpha-helices flanking a beta strand. The SLH sequences are fairly divergent with an average identity of about 25%. It is however possible to build a sequence pattern that starts at the second position of the domain and that spans 3/4 of its length.

Consensus pattern[LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[FYWPD]-x(4)-[LIV]-x(2)-[GTALV]-x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x- [PGAV]-x(3,10)-[LIVMA]-[STKR]-[RY]-x-[EQ]-x-[STALIVM] Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT/NONE.

[1] Beveridge T.J. Curr. Opin. Struct. Biol. 4:204-212(1994).

[2] Lupas A., Engelhardt H., Peters J., Santarius U., Volker S., Baumeister W. J. Bacteriol. 176:1224-1233(1994).

[3] Lemaire M., Ohayon H., Gounon P., Fujino T., Beguin P. J. Bacteriol. 177:2451-2459(1995).

903. Queuine tRNA-ribosyltransferase (TGT)

This is a family of queuine tRNA-ribosyltransferases EC:2.4.2.29, also known as tRNA-guanine transglycosylase and guanine insertion enzyme. Queuine tRNA-ribosyltransferase modifies tRNAs for asparagine, aspartic acid, histidine and tyrosine with queuine. It catalyses the exchange of guanine-34 at the wobble position with 7-aminomethyl-7-deazaguanine, and the addition of a cyclopentenediol moiety to 7-aminomethyl-7-deazaguanine-34 tRNA; giving a hypermodified base queuine in the wobble position [1,2]. The aligned region contains a zinc binding motif C-x-C-x2-C-x29-H, and important tRNA and 7-aminomethyl-7-deazaguanine binding residues [1]. Number of members: 27

[1] Romier C, Reuter K, Suck D, Ficner R; Medline: 96256303 "Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange." *EMBO J* 1996;15:2850-2857.

[2] Garcia GA, Koch KA, Chong S; Medline: 93287116 "tRNA-guanine transglycosylase from *Escherichia coli*. Overexpression, purification and quaternary structure." *J Mol Biol* 1993;231:489-497.

904. ThiC Family (ThiC)

ThiC is found within the thiamine biosynthesis operon. ThiC is involved in pyrimidine biosynthesis [2]. ThiC catalyzes the substitution of the pyrophosphate of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by 4-methyl-5-(beta-hydroxyethyl)thiazole phosphate to yield thiamine phosphate [3]. Number of members: 12

[1] Vander Horn PB, Backstrom AD, Stewart V, Begley TP; Medline: 93163063 "Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in *Escherichia coli* K-12." *J Bacteriol* 1993;175:982-992.

[2] Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon AP, Taylor S, Campobasso N, Chiu HJ, Kinsland C, Reddick JJ, Xi J; Medline: 99311269 "Thiamin biosynthesis in prokaryotes." *Arch Microbiol* 1999;171:293-300.

[3] Zhang Y, Taylor SV, Chiu HJ, Begley TP; Medline: 97284509 "Characterization of the *Bacillus subtilis* thiC operon involved in thiamine biosynthesis." *J Bacteriol* 1997;179:3030-3035.

905. Putative tRNA binding domain (tRNA_bind)

This domain is found in prokaryotic methionyl-tRNA synthetases, prokaryotic phenylalanyl tRNA synthetases the yeast GU4 nucleic-binding protein (G4p1 or p42, ARC1) [2], human tyrosyl-tRNA synthetase [1], and endothelial-monocyte activating polypeptide II. G4p1 binds specifically to tRNA form a complex with methionyl-tRNA synthetases [2]. In human tyrosyl-tRNA synthetase this domain may direct tRNA to the active site of the enzyme [2]. This domain may perform a common function in tRNA aminoacylation [1]. Number of members: 12

[1] Kleeman TA, Wei D, Simpson KL, First EA; Medline: 97306356 "Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine." *J Biol Chem* 1997;272:14420-14425.

- 5 [2] Simos G, Segref A, Fasiolo F, Hellmuth K, Shevchenko A, Mann M, Hurt EC; Medline: 97050848 "The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases." *EMBO J* 1996;15:5437-5448.

10 906. UbiA prenyltransferase family signature (UbiA)

The following prenyltransferases are evolutionary related [1,2]:

- Bacterial 4-hydroxybenzoate octaprenyltransferase (gene *ubiA*).
 - Yeast mitochondrial para-hydroxybenzoate--polyprenyltransferase (gene *COQ2*).
 - Protoheme IX farnesyltransferase (heme O synthase) from yeast and mammals (gene
- 15 COX10) and from bacteria (genes *cyoE* or *ctaB*).

These proteins probably contain seven transmembrane segments. The best conserved region is located in a loop between the second and third of these segments and was used as a signature pattern.

20 Consensus pattern N-x(3)-[DE]-x(2)-[LIF]-D-x(2)-[VM]-x-R-[ST]-x(2)-R-x(4)-G Sequences known to belong to this class detected by the pattern ALL. Other sequence(s) detected in SWISS-PROT/NONE.

- 25 [1] Melzer M., Heide L. *Biochim. Biophys. Acta* 1212:93-102(1994).
[2] Mogi T., Saiki K., Anraku Y. *Mol. Microbiol.* 14:391-398(1994).

907. Uncharacterized protein family UPF0044 signature (UPF0044)

30 The following uncharacterized proteins have been shown [1] to be highly similar:

- *Bacillus subtilis* hypothetical protein *yqeL*.
- *Escherichia coli* hypothetical protein *yhbY* and HI1333, the corresponding *Haemophilus influenzae* protein.

736

- *Methanococcus jannaschii* hypothetical protein MJ0652.

These are small proteins of 10 to 15 Kd. They can be picked up in the database by the following pattern. This pattern is located in the N-terminal part of these proteins.

- 5 Consensus pattern L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

908. ATP synthase (C/AC39) subunit (vATP-synt_AC39)

- 10 This family includes the AC39 subunit from vacuolar ATP synthase Swiss:P32366 [1], and the C subunit from archaebacterial ATP synthase [2]. The family also includes subunit C from the Sodium transporting ATP synthase from *Enterococcus hirae* Swiss:P43456 [3].
Number of members: 12

- 15 [1] Bauerle C, Ho MN, Lindorfer MA, Stevens TH; Medline: 93286119 "The *Saccharomyces cerevisiae* VMA6 gene encodes the 36-kDa subunit of the vacuolar H(+)-ATPase membrane sector." J Biol Chem 1993;268:12749-12757.
[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968 "Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon
20 *Methanosarcina mazei* Go1." J Biol Chem 1996;271:18843-18852.
[3] Takase K, Kakinuma S, Yamato I, Konishi K, Igarashi K, Kakinuma Y; Medline: 94209269 "Sequencing and characterization of the ntp gene cluster for vacuolar- type Na(+)-translocating ATPase of *Enterococcus hirae*." J Biol Chem 1994;269:11037-11044.

25

909. ATP synthase (E/31 kDa) subunit (vATP-synt_E)

This family includes the vacuolar ATP synthase E subunit [1], as well as the archaebacterial ATP synthase E subunit [2]. Number of members: 24

- 30 [1] Foury F; Medline: 91009356 "The 31-kDa polypeptide is an essential subunit of the vacuolar ATPase in *Saccharomyces cerevisiae*." J Biol Chem 1990;265:18554-18560.

[2] Wilms R, Freiberg C, Wegerle E, Meier I, Mayer F, Muller V; Medline: 96324968
 "Subunit structure and organization of the genes of the A1A0 ATPase from the Archaeon
 Methanosarcina mazei Go1." J Biol Chem 1996;271:18843-18852.

910. (WW)

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

- Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.
- Utrophin, a dystrophin-like protein of unknown function.
- Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].
- Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].

- Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>, followed by a histidine-rich region, 3 WW domains and a HECT domain.

- Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

- Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in Drosophila and in mammals (gene PIN1).

- Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.

- IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

- Yeast pre-mRNA processing protein PRP40, Caenorhabditis elegans ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2- type myosin, each containing two WW-domains at the N-terminus.

- Caenorhabditis elegans hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.

- Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P
Sequences known to belong to this class detected by the pattern ALL. Other sequence(s)
detected in SWISS-PROT8. Sequences known to belong to this class detected by the
profile ALL.

[1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).

[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).

[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).

[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).

[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).

[6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).

911. Xeroderma pigmentosum (XP) [1] (XPG_1)

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. The defect in XP-G can be corrected by a 133 Kd nuclear protein called XPG (or XPGC) [2].

XPG belongs to a family of proteins [2,3,4,5,6] that are composed of two main subsets:

- Subset 1, to which belongs XPG, RAD2 from budding yeast and rad13 from fission yeast. RAD2 and XPG are single-stranded DNA endonucleases [7,8]. XPG makes the 3'incision in human DNA nucleotide excision repair [9].
- Subset 2, to which belongs mouse and human FEN-1, rad2 from fission yeast, and RAD27 from budding yeast. FEN-1 is a structure-specific endonuclease.

In addition to the proteins listed in the above groups, this family also includes:

- Fission yeast *exo1*, a 5'→3' double-stranded DNA exonuclease that could act in a pathway that corrects mismatched base pairs.
- Yeast *EXO1* (*DHS1*), a protein with probably the same function as *exo1*.
- Yeast *DIN7*.

Sequence alignment of this family of proteins reveals that similarities are largely confined to two regions. The first is located at the N-terminal extremity (N-region) and corresponds to the first 95 to 105 amino acids. The second region is internal (I-region) and found towards the C-terminus; it spans about 140 residues and contains a highly conserved core of 27 amino acids that includes a conserved pentapeptide (E-A-[DE]-A-[QS]). It is possible that the conserved acidic residues are involved in the catalytic mechanism of DNA excision repair in XPG. The amino acids linking the N- and I-regions are not conserved; indeed, they are largely absent from proteins belonging to the second subset.

Two signature patterns were developed for these proteins. The first corresponds to the central part of the N-region, the second to part of the I-region and includes the putative catalytic core pentapeptide.

Consensus pattern [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

Consensus pattern [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]- [QS]-[CLM] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

[2] Scherly D., Nusspikel T., Corlet J., Ucla C., Bairoch A., Clarkson S.G. Nature 363:182-185(1993).

[3] Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z., Lehmann A.R. Nucleic Acids Res. 21:1345-1349(1993).

[4] Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S., Lehmann A.R., Carr A.M., Watts F.Z. Mol. Cell. Biol. 14:4878-4888(1994).

[5] Harrington J.J., Lieber M.R. Genes Dev. 8:1344-1355(1994).

[6] Szankasi P., Smith G.R. Science 267:1166-1169(1995).

[7] Habraken Y., Sung P., Prakash L., Prakash S. Nature 366:365-368(1993).

[8] O'Donovan A., Scherly D., Clarkson S.G., Wood R.D. J. Biol. Chem. 269:15965-15968(1994).

[9] O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D. Nature 371:432-435(1994).

912. 5-formyltetrahydrofolate cyclo-ligase (5-FTHF_cyc-lig)

5-formyltetrahydrofolate cyclo-ligase or methenyl-THF synthetase EC:6.3.3.2 catalyses the interchange of 5-formyltetrahydrofolate (5-FTHF) to 5-10-methenyltetrahydrofolate, this

requires ATP and Mg^{2+} [1]. 5-FTHF is used in chemotherapy where it is clinically known as Leucovorin [2].

Number of members: 23

[1] Dayan A, Bertrand R, Beauchemin M, Chahla D, Mamo A, Filion M, Skup D, Massie B, Jolivet J; Medline: 96096540 "Cloning and characterization of the human 5,10-methenyltetrahydrofolate synthetase-encoding cDNA." *Gene* 1995;165:307-311.

[2] Maras B, Stover P, Valiante S, Barra D, Schirch V; Medline: 94308074 "Primary structure and tetrahydropteroylglutamate binding site of rabbit liver cytosolic 5,10-methenyltetrahydrofolate synthetase." *J Biol Chem* 1994;269:18429-18433.

913. Cytosolic long-chain acyl-CoA thioester hydrolase (Acyl-CoA_hydro)

This family consist of various cytosolic long-chain acyl-CoA thioester hydrolases including human and rat [1,2]. The aligned region is repeated with in the sequence of human and rat cytosolic long-chain acyl-CoA thioester hydrolases of this family. Long-chain acyl-CoA hydrolases hydrolyse palmitoyl-CoA to CoA and palmitate, they also catalyse the hydrolysis of other long chain fatty acyl-CoA thioesters. Long-chain acyl-CoA hydrolases are present in all living organisms and they may provide a mechanism for the control of lipid metabolism [1].

Number of members: 24

[1]Yamada J, Furihata T, Iida N, Watanabe T, Hosokawa M, Satoh T, Someya A, Nagaoka I, Suga T; Medline: 97236308 "Molecular cloning and expression of cDNAs encoding rat brain and liver cytosolic long-chain acyl-CoA hydrolases." *Biochem Biophys Res Commun* 1997;232:198-203.

[2] Broustas CG, Larkins LK, Uhler MD, Hajra AK; Medline: 96209964 "Molecular cloning and expression of cDNA encoding rat brain cytosolic acyl-coenzyme A thioester hydrolase." *J Biol Chem* 1996;271:10470-10476.

914. Agglutinin

Lectin (probable mannose binding)

Members of this family are plant lectins. Many if not all are mannose specific.

Number of members: 87

[1] Wright CS, Hester G; Medline: 97094989 "The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes." Structure 1996;4:1339-1352.

915. (ANF_RECEPTORS)

Natriuretic peptides are hormones involved in the regulation of fluid and electrolyte homeostasis. These hormones stimulate the intracellular production of cyclic GMP as a second messenger.

Currently, three types of natriuretic peptide receptors are known [1,2]. Two express guanylate cyclase activity: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP. The third receptor (ANP-C) is probably responsible for the clearance of ANP from the circulation and does not play a role in signal transduction.

GC-A and GC-B are plasma membrane-bound proteins that share the following topology: an N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain followed by a large cytoplasmic C- terminal region that can be subdivided into two domains: a protein kinase-like domain (see <PDOC00100>) that appears important for proper signalling and a guanylate cyclase catalytic domain (see <PDOC00425>). The topology of ANP-C is different: like GC-A and -B it possesses an extracellular ligand-binding region and a transmembrane domain, but its cytoplasmic domain is very short.

A pattern was developed from the ligand-binding region of natriuretic peptide receptors based on a highly conserved region located in the N-terminal part of the domain.

Consensus patternG-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT/NONE.

[1] Garbers D.L. New Biol. 2:499-504(1990).

[2] Schulz S., Chinkers M., Garbers D.L. FASEB J. 2:2026-2035(1989).

5 916. (Apocytochrome)

Cytochrome c family heme-binding site signature

In proteins belonging to cytochrome c family [1], the heme group is covalently attached by thioether bonds to two conserved cysteine residues. The consensus sequence for this site is Cys-X-X-Cys-His and the histidine residue is one of the two axial ligands of the heme iron. This arrangement is shared by all proteins known to belong to cytochrome c family, which presently includes cytochromes c, c', c1 to c6, c550 to c556, cc3/Hmc, cytochrome f and reaction center cytochrome c.

Consensus pattern C-{CPWHF}-{CPWR}-C-H-{CFYW} Sequences known to belong to this class detected by the pattern ALL, except for four cytochrome c's which lack the first thioether bond. Other sequence(s) detected in SWISS-PROT454.

Note: some cytochrome c's have more than a single bound heme group c4 has 2, c7 has 3, c3 has 4, the reaction center has 4, and cc3/Hmc has 16 !

[1] Mathews F.S. Prog. Biophys. Mol. Biol. 45:1-56(1985).

917. ATP-synt_A-c. ATP synthase Alpha chain, C terminal

[1] Medline: 94344236. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Abrahams JP, Leslie AG, Lutter R, Walker JE; Nature 1994;370:621-628.

Number of members: 125

918. (Basic)

Myc-type, 'helix-loop-helix' dimerization domain signature

HELIX_LOOP_HELIX

A number of eukaryotic proteins, which probably are sequence specific DNA- binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed [1] that this domain is formed of two amphipathic helices joined by a variable length linker region that could form a loop. This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in the proteins listed below [2,3,E1,E2]. Most of these proteins have an extra basic region of about 15 amino acid residues that is adjacent to the HLH domain and specifically binds to DNA. They are referred as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). Members of the bHLH family bind variations on the core sequence 'CANNTG', also referred to as the E-box motif. The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW.

- The myc family of cellular oncogenes [4], which is currently known to contain four members: c-myc [E3], N-myc, L-myc, and B-myc. The myc genes are thought to play a role in cellular differentiation and proliferation.
- Proteins involved in myogenesis (the induction of muscle cells). In mammals MyoD1 (Myf-3), myogenin (Myf-4), Myf-5, and Myf-6 (Mrf4 or herculin), in birds CMD1 (QMF-1), in *Xenopus* MyoD and MF25, in *Caenorhabditis elegans* CeMyoD, and in *Drosophila nautilus* (nau).
- Vertebrate proteins that bind specific DNA sequences ('E boxes') in various immunoglobulin chains enhancers: E2A or ITF-1 (E12/pan-2 and E47/pan-1), ITF-2 (tcf4), TFE3, and TFEF.
- Vertebrate neurogenic differentiation factor 1 that acts as differentiation factor during neurogenesis.
- Vertebrate MAX protein, a transcription regulator that forms a sequence- specific DNA-binding protein complex with myc or mad.
- Vertebrate Max Interacting Protein 1 (MXI1 protein) which acts as a transcriptional repressor and may antagonize myc transcriptional activity by competing for max.

- Proteins of the bHLH/PAS superfamily which are transcriptional activators. In mammals, AH receptor nuclear translocator (ARNT), single-minded homologs (SIM1 and SIM2), hypoxia-inducible factor 1 alpha (HIF1A), AH receptor (AHR), neuronal pas domain proteins (NPAS1 and NPAS2), endothelial pas domain protein 1 (EPAS1), mouse ARNT2, and human BMAL1. In drosophila, single-minded (SIM), AH receptor nuclear translocator (ARNT), trachealess protein (TRH), and similar protein (SIMA).

- Mammalian transcription factors HES, which repress transcription by acting on two types of DNA sequences, the E box and the N box.

- Mammalian MAD protein (max dimerizer) which acts as transcriptional repressor and may antagonize myc transcriptional activity by competing for max.

- Mammalian Upstream Stimulatory Factor 1 and 2 (USF1 and USF2), which bind to a symmetrical DNA sequence that is found in a variety of viral and cellular promoters.

- Human lyl-1 protein; which is involved, by chromosomal translocation, in T- cell leukemia.

- Human transcription factor AP-4.

- Mouse helix-loop-helix proteins MATH-1 and MATH-2 which activate E box- dependent transcription in collaboration with E47.

- Mammalian stem cell protein (SCL) (also known as tal1), a protein which may play an important role in hemopoietic differentiation. SCL is involved, by chromosomal translocation, in stem-cell leukemia.

- Mammalian proteins Id1 to Id4 [5]. Id (inhibitor of DNA binding) proteins lack a basic DNA-binding domain but are able to form heterodimers with other HLH proteins, thereby inhibiting binding to DNA.

- Drosophila extra-macrochaetae (emc) protein, which participates in sensory organ patterning by antagonizing the neurogenic activity of the achaete- scute complex. Emc is the homolog of mammalian Id proteins.

- Human Sterol Regulatory Element Binding Protein 1 (SREBP-1), a transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) found in the flanking region of the LDLR gene and in other genes.

- Drosophila achaete-scute (AS-C) complex proteins T3 (l'sc), T4 (scute), T5 (achaete) and T8 (asense). The AS-C proteins are involved in the determination of the neuronal precursors in the peripheral nervous system and the central nervous system.

- Mammalian homologs of achaete-scute proteins, the MASH-1 and MASH-2 proteins.

- Drosophila atonal protein (ato) which is involved in neurogenesis.

- *Drosophila* daughterless (da) protein, which is essential for neurogenesis and sex-determination.

- *Drosophila* deadpan (dpn), a hairy-like protein involved in the functional differentiation of neurons.

5 - *Drosophila* delilah (dei) protein, which is plays an important role in the differentiation of epidermal cells into muscle.

- *Drosophila* hairy (h) protein, a transcriptional repressor which regulates the embryonic segmentation and adult bristle patterning.

10 - *Drosophila* enhancer of split proteins E(spl), that are hairy-like proteins active during neurogenesis. also act as transcriptional repressors.

- *Drosophila* twist (twi) protein, which is involved in the establishment of germ layers in embryos.

- Maize anthocyanin regulatory proteins R-S and LC.

15 - Yeast centromere-binding protein 1 (CPF1 or CBF1). This protein is involved in chromosomal segregation. It binds to a highly conserved DNA sequence, found in centromeres and in several promoters.

- Yeast INO2 and INO4 proteins.

- Yeast phosphate system positive regulatory protein PHO4 which interacts with the upstream activating sequence of several acid phosphatase genes.

20 - Yeast serine-rich protein TYE7 that is required for ty-mediated ADH2 expression.

- *Neurospora crassa* nuc-1, a protein that activates the transcription of structural genes for phosphorus acquisition.

- Fission yeast protein esc1 which is involved in the sexual differentiation process.

25 The schematic representation of the helix-loop-helix domain is shown here:

XXXXXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXXXXXXXXXXXX Amphipathic
helix 1 Loop Amphipathic helix 2

30 The signature pattern that had been developed to detect this domain spans completely the second amphipathic helix.

Consensus pattern[DENSTAP]-[KR]-[LIVMAGSNT]-{FYWCPhKR}-[LIVMT]-[LIVM]-
x(2)-{STAV}-[LIVMSTACKR]-x-[VMFYH]-[LIVMTA]-{P}-{P}-[LIVMRKHQ]

Sequences known to belong to this class detected by the pattern the majority but far from all.
Other sequence(s) detected in SWISS-PROT135.

[1] Murre C., McCaw P.S., Baltimore D. Cell 56:777-783(1989).

5 [2] Garrel J., Campuzano S. BioEssays 13:493-498(1991).

[3] Kato G.J., Dang C.V. FASEB J. 6:3065-3072(1992).

[4] Krause M., Fire A., Harrison S.W., Priess J., Weintraub H. Cell 63:907-919(1990).

[5] Riechmann V., van Cruechten I., Sablitzky F. Nucleic Acids Res. 22:749-755(1994).

10 919. (Beta-lactamase)

Beta-lactamases classes -A, -C, and -D active site

Beta-lactamases (EC 3.5.2.6) [1,2] are enzymes which catalyze the hydrolysis of an amide bond in the beta-lactam ring of antibiotics belonging to the penicillin/cephalosporin family. Four kinds of beta-lactamase have been identified [3]. Class-B enzymes are zinc containing proteins whilst class -A, C and D enzymes are serine hydrolases. The three classes of serine beta-lactamases are evolutionary related and belong to a superfamily [4] that also includes DD-peptidases and a variety of other penicillin-binding proteins (PBP's). All these proteins contain a Ser-x-x-Lys motif, where the serine is the active site residue. Although clearly homologous, the sequences of the three classes of serine beta-lactamases exhibit a large degree of variability and only a small number of residues are conserved in addition to the catalytic serine.

25 Since a pattern detecting all serine beta-lactamases would also pick up many unrelated sequences, it was decided to provide specific patterns, centered on the active site serine, for each of the three classes.

Consensus pattern [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC] [S is the active site residue] Sequences known to belong to this class detected by the patternALL class-A beta-lactamases. Other sequence(s) detected in SWISS-PROT7.

30

Consensus pattern F-E-[LIVM]-G-S-[LIVMG]-[SA]-K [The first S is the active site residue]
 Sequences known to belong to this class detected by the patternALL class-C beta-lactamases.
 Other sequence(s) detected in SWISS-PROT/NONE.

- 5 Consensus pattern [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI] [S is the active site residue]
 Sequences known to belong to this class detected by the patternALL class-D beta-lactamases. Other sequence(s) detected in SWISS-PROT/NONE.

[1] Ambler R.P. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 289:321-331(1980).

- 10 [2] Pastor N., Pinero D., Valdes A.M., Soberon X. Mol. Microbiol. 4:1957-1965(1990).

[3] Bush K. Antimicrob. Agents Chemother. 33:259-263(1989).

[4] Joris B., Ghuysen J.-M., Dive G., Renard A., Dideberg O., Charlier P., Frere J.M., Kelly J.A., Boyington J.C., Moews P.C., Knox J.R. Biochem. J. 250:313-324(1988).

- 15 920. Biotin protein ligase (BPL)

Biotin is covalently attached at the active site of certain enzymes that transfer carbon dioxide from bicarbonate to organic acids to form cellular metabolites. Biotin protein ligase (BPL) is the enzyme responsible for attaching biotin to a specific lysine at the active site of biotin
 20 enzymes. Each organism probably has only one BPL. Biotin attachment is a two step reaction that results in the formation of an amide linkage between the carboxyl group of biotin and the epsilon-amino group of the modified lysine [2].

Number of members: 26

- 25 [1] Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW; Medline: 93028443
 "Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains." Proc Natl Acad Sci USA 1992;89:9257-9261.

- [2] Chapman-Smith A, Cronan JE Jr; Medline: 10470036 "The enzymatic biotinylation of
 30 proteins: a post-translational modification of exceptional specificity." Trends Biochem Sci 1999;24:359-363.

921. (BRCA2_repeat)

The alignment covers only the most conserved region of the repeat. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature

[1] Bork P, Blomberg N, Nilges M; Medline: 96241568 "Internal repeats in the BRCA2 protein sequence." Nat Genet 1996;13:22-23.

Number of members: 63

922. (C6)

This domain of unknown function is found in the *C. elegans* protein Swiss:Q19522. It is presumed to be an extracellular domain. The C6 domain contains six conserved cysteine residues in most copies of the domain. However some copies of the domain are missing cysteine residues 1 and 3 suggesting that these form a disulphide bridge.

Number of members: 23

923. Cadherin cytoplasmic region (Cadherin_C_term)

Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn [1].

Number of members: 59

[1] Yap AS, Niessen CM, Gumbiner BM; Medline: 98234411 "The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with p120ctn." J Cell Biol 1998;141:779-789.

[2] Barth AI, Nathke IS, Nelson WJ; Medline: 97471931 "Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways." Curr Opin Cell Biol 1997;9:683-690.

[3] Braga VM, Machesky LM, Hall A, Hotchin NA; Medline: 97327766 "The small GTPases Rho and Rac are required for the establishment of cadherin-dependent cell-cell contacts." J Cell Biol 1997;137:1421-1431.

5 924. Clathrin propeller repeat (Clathrin_propel)

Clathrin is the scaffold protein of the basket-like coat that surrounds coated vesicles. The soluble assembly unit, a triskelion, contains three heavy chains and three light chains in an extended three-legged structure. Each leg contains one heavy and one light chain. The N-terminus of the heavy chain is known as the globular domain, and is composed of seven repeats which form a beta propeller [1].

Number of members: 61

[1] ter Haar E, Musacchio A, Harrison SC, Kirchhausen T; Medline: 99043510 "Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker." Cell. 1998;95:563-573.

925. Respiratory-chain NADH dehydrogenase 30 Kd subunit signature (complex1_30Kd)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 30 Kd (in mammals) which has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in *Neurospora crassa*.
- Mitochondrial encoded in *Paramecium* (protein P1), and in the slime mold *Dictyostelium discoideum* (ORF 209).
- Chloroplast encoded in various higher plants (ORF 159). It is also present in bacteria:
 - In the cyanobacteria *Synechocystis* strain PCC 6803 (gene *ndhJ*).
 - Subunit C of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoC*).
 - Subunit NQO5 of *Paracoccus denitrificans* NADH-ubiquinone oxidoreductase.

This protein, in its mature form, consists of from 157 to 266 amino acid residues. The best conserved region is located in the C-terminal section and can be used as a signature pattern.

- 5 Consensus pattern E-R-E-x(2)-[DE]-[LIVMFY](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMYS] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROTNONE.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

- 10 [2]Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

926. Respiratory-chain NADH dehydrogenase 49 Kd subunit signature (complex1_49Kd)

- 15 Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there is one with a molecular weight of 49 Kd (in mammals), which is the third largest subunit of complex I and is a component of the iron-sulfur (IP) fragment of the enzyme. It seems to bind a 4Fe-4S iron-sulfur cluster. The 49 Kd subunit has been found to be:

- Nuclear encoded, as a precursor form with a transit peptide in mammals, and in *Neurospora crassa*.
 - Mitochondrial encoded in protozoan such as *Paramecium* (ORF 400), *Leishmania* and *Trypanosoma* (MURF 3).
 - Chloroplast encoded in various higher plants (ORF 392).
- The 49 Kd subunit is highly similar to [3,4]:
- Subunit D of *Escherichia coli* NADH-ubiquinone oxidoreductase (gene *nuoD*).
 - Subunit NQO4 of *Paracoccus denitrificans* NADH-ubiquinone oxidoreductase.
 - 30 - Subunit 5 of *Escherichia coli* formate hydrogenlyase (gene *hycE*).
 - Subunit G of *Escherichia coli* hydrogenase-4 (gene *hyfG*).

A highly conserved region was selected as signature pattern, located in the N-terminal section of this subunit.

Consensus pattern [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMTN]-x-E-x-[KRO] Sequences known to belong to this class detected by the patternALL.

- 5 [1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).
- [2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).
- [3] Fearnley I.M., Walker J.E. Biochim. Biophys. Acta 1140:105-134(1992).
- [4] Weidner U., Geier S., Ptöck A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

10

927. (COX2)

Cytochrome c oxidase (EC 1.9.3.1) [1,2] is an oligomeric enzymatic complex which is a component of the respiratory chain and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. The enzyme complex consists of 3-4 subunits (prokaryotes) to up to 13 polypeptides (mammals).

15

Subunit 2 (CO II) transfers the electrons from cytochrome c to the catalytic subunit 1. It contains two adjacent transmembrane regions in its N-terminus and the major part of the protein is exposed to the periplasmic or to the mitochondrial intermembrane space, respectively. CO II provides the substrate-binding site and contains a copper center called Cu(A), probably the primary acceptor in cytochrome c oxidase. An exception is the corresponding subunit of the cbb3-type oxidase which lacks the copper A redox-center.

20

Several bacterial CO II have a C-terminal extension that contains a covalently bound heme c.

25

It has been shown [3,4] that nitrous oxide reductase (EC 1.7.99.6) (gene nosZ) of *Pseudomonas* has sequence similarity in its C-terminus to CO II. This enzyme is part of the bacterial respiratory system which is activated under anaerobic conditions in the presence of nitrate or nitrous oxide. NosZ is a periplasmic homodimer that contains a dinuclear copper center, probably located in a 3-dimensional fold similar to the cupredoxin-like fold that has been suggested for the copper-binding site of CO II [3].

30

The dinuclear purple copper center is formed by 2 histidines and 2 cysteines [5]. This region was used as a signature pattern. The conserved valine and the conserved methionine are said to be involved in stabilizing the copper-binding fold by interacting with each other.

- 5 Consensus pattern V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M [The two C's and two H's are copper ligands] Sequences known to belong to this class detected by the patternALL, except for *Paramecium primaurelia* as well as in some plants where the pattern ends with Thr; an RNA editing event at this position could change this Thr to Met.

- 10 Note: cytochrome cbb(3) subunit 2 does not belong to this family.

[1] Capaldi R.A., Malatesta F., Darley-Usmar V.M. *Biochim. Biophys. Acta* 726:135-148(1983).

[2] Garcia-Horsman J.A., Barquera B., Rumbley J., Ma J., Gennis R.B. *J. Bacteriol.* 176:5587-5600(1994).

[3] van der Oost J., Lappalainen P., Musacchio A., Warne A., Lemieux L., Rumbley J., Gennis R.B., Aasa R., Pascher T., Malmstrom B.G., Saraste M. *EMBO J.* 11:3209-3217(1992).

[4] Zumft W.G., Dreutsch A., Loechele S., Cuypers H., Friedrich B., Schneider B. *Eur. J. Biochem.* 208:31-40(1992).

928. Cytochrome C assembly protein (CytC_asm)

This family consists of various proteins involved in cytochrome c assembly from mitochondria and bacteria; CycK from *Rhizobium*[3], CcmC from *E. coli* and *Paracoccus denitrificans* [2,1] and orf240 from wheat mitochondria [4]. The members of this family are probably integral membrane proteins with six predicted transmembrane helices. It has been proposed that members of this family comprise a membrane component of an ABC (ATP binding cassette) transporter complex. It is also proposed that this transporter is necessary for transport of some component needed for cytochrome c assembly. One member CycK contains a putative heme-binding motif [3], orf240 also contains a putative heme-binding motif and is a proposed ABC transporter with c-type heme as its proposed substrate [4].

However it seems unlikely that all members of this family transport heme nor c-type apocytochromes because CcmC in the putative CcmABC transporter transports neither [1].

Number of members: 67

- 5 [1] Page D, Pearce DA, Norris HA, Ferguson SJ; Medline: 97195802 "The *Paracoccus denitrificans* ccmA, B and C genes: cloning and sequencing, and analysis of the potential of their products to form a haem or apo-c-type cytochrome transporter. MICROBIOLOGY 1997;143:563-576.
- 10 [2] Thoeny-meyer L, Fischer F, Kunzler P, Ritz D, Hennecke H; Medline: 95362656 "Escherichia coli genes required for cytochrome c maturation." J. BACTERIOL 1995;177:4321-4326.
- 15 [3] Delgado MJ, Yeoman KH, Wu G, Vargas C, Davies A, Poole RK, Johnston AWB, Downie JA; Medline: 95394794 "Characterization of the cychJKL genes involved in cytochrome c biogenesis and symbiotic nitrogen fixation in *Rhizobium leguminosarum*." J. BACTERIOL 1995;177:4927-4934.
- [4] Bonnard G, Grienemberger JM; Medline: 95124303 "A gene proposed to encode a transmembrane domain of an ABC transporter is expressed in wheat mitochondria." MOL. GEN. GENET 1995;246:91-99.
- 20 929. Cytochrome b559 subunits heme-binding site signature (cytochr_b559)

Cytochrome b559 [1] is an essential component of photosystem II complex from oxygenic photosynthetic organisms. It is an integral thylakoid membrane protein composed of two subunits, alpha (gene psbE) and beta (gene psbF), each of which contains a histidine residue located in a transmembrane region. The two histidines coordinate the heme iron of cytochrome b559.

The region around the heme-binding residue of both subunits is very similar and can be used as a signature pattern.

Consensus pattern[LIV]-x-[ST]-[LIVF]-R-[FYW]-x(2)-[IV]-H-[STGA]-[LIV]-[STGA]-[IV]-P [H is the heme iron ligand] Sequences known to belong to this class detected by the patternALL. Other sequence(s) detected in SWISS-PROT NONE.

[1] Pakrasi H.B., de Ciechi P., Whitmarsh J. EMBO J. 10:1619-1627(1991).

930. Cytochrome b/b6 signatures (Cytochrome_b)

In the mitochondrion of eukaryotes and in aerobic prokaryotes, cytochrome b is a component of respiratory chain complex III (EC 1.10.2.2) - also known as the bc1 complex or ubiquinol-cytochrome c reductase. In plant chloroplasts and cyanobacteria, there is an analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase (EC 1.10.99.1), also known as the b6f complex.

Cytochrome b/b6 [1,2] is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the petB and petD genes. The sequence of petB is colinear with the N-terminal part of mitochondrial cytochrome b, while petD corresponds to the C-terminal part. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups.

Apart from regions around some of the histidine heme ligands, there are a few conserved regions in the sequence of b/b6. The best conserved of these regions includes an invariant P-E-W triplet which lies in the loop that separates the fifth and sixth transmembrane segments. It seems to be important for electron transfer at the ubiquinone redox site - called Qz or Qo (where o stands for outside) - located on the outer side of the membrane.

A schematic representation of the structure of cytochrome b/b6 is shown below.

```
+---Fe-b562----+ | +---Fe-b566--|+ |||
xxxxxxxxxxHxHxxxxxxxxxxHxHxxxxxxxxxxPEWxxxxxxxxxxxxxxxxxx <-----
---Cytochrome-b-----> <---Cytochrome-b6-petB-----> <---Cytochrome-
b6-petD----->
```

Two signature patterns were developed for cytochrome b/b6. The first includes the first conserved histidine of b/b6, which is a heme b562 ligand; the second includes the conserved PEW triplet.

- 5 Consensus pattern [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H [H is a heme b562 ligand] Sequences known to belong to this class detected by the patternALL, except for 5 sequences.

- 10 Consensus pattern P-[DE]-W-[FY]-[LFY](2) Sequences known to belong to this class detected by the patternALL, except for *Odocoileus hemionus* (mule deer) and *Paramecium tetraurelia* cytochrome b.

[1] Howell N. J. Mol. Evol. 29:157-169(1989).

- 15 [2] Esposti M.D., de Vries S., Crimi M., Ghelli A., Patarnello T., Meyer A. Biochim. Biophys. Acta 1143:243-271(1993).

931. Phorbol esters / diacylglycerol binding domain (DAG_PE-bind)

- 20 Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) [1]. Phorbol esters can directly stimulate PKC. The N- terminal region of PKC, known as C1, has been shown [2] to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two
- 25 copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in the following proteins:

- 30 - Diacylglycerol kinase (EC 2.7.1.107) (DGK) [3], the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals.
- N-chimaerin. A brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown [4,5] to be able to bind phorbol esters.

- The raf/mil family of serine/threonine protein kinases. These protein kinases contain a single N-terminal copy of the DAG/PE-binding domain.

- The unc-13 protein from *Caenorhabditis elegans*. Its function is not known but it contains a copy of the DAG/PE-binding domain in its central section and has been shown to bind specifically to a phorbol ester in the presence of calcium [6].

- The vav oncogene. Vav was generated by a genetic rearrangement during gene transfer assays. Its expression seems to be restricted to cells of hematopoietic origin. Vav seems to contain a DAG/PE-binding domain in the central part of the protein.

- The *Drosophila* GTPase activating protein rotund.

The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. A signature pattern was developed that spans completely the DAG/PE domain.

Consensus pattern H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HID]-x(2)-C-x(5,9)-C [All the C and H are involved in binding Zinc] Sequences known to belong to this class detected by the pattern ALL, except a few DGK's.

[1] Azzi A., Boscoboinik D., Hensey C. *Eur. J. Biochem.* 208:547-557(1992).

[2] Ono Y., Fujii T., Igarashi K., Kuno T., Tanaka C., Kikkawa U., Nishizuka Y. *Proc. Natl. Acad. Sci. U.S.A.* 86:4868-4871(1989).

[3] Sakane F., Yamada K., Kanoh H., Yokoyama C., Tanabe T. *Nature* 344:345-348(1990).

[4] Ahmed S., Kozma R., Monfries C., Hall C., Lim H.H., Smith P., Lim L. *Biochem. J.* 272:767-773(1990).

[5] Ahmed S., Kozma R., Lee J., Monfries C., Harden N., Lim L. *Biochem. J.* 280:233-241(1991).

[6] Ahmed S., Maruyama I.N., Kozma R., Lee J., Brenner S., Lim L. *Biochem. J.* 287:995-999(1992).

[7] Boguski M.S., Bairoch A., Attwood T.K., Michaels G.S. *Nature* 358:113-113(1992).

932. 3-dehydroquinate synthase (DHQ_synthase)

[1] Barten R, Meyer TF; Medline: 98273626 "Cloning and characterisation of the *Neisseria gonorrhoeae* *aroB* gene." *Mol Gen Genet* 1998;258:34-44.

[2] Hawkins AR, Lamb HK; Medline: 96048023 "The molecular biology of multidomain proteins. Selected examples." *Eur J Biochem* 1995;232:7-18.

5

The 3-dehydroquinate synthase EC:4.6.1.3 domain is present in isolation in various bacterial 3-dehydroquinate synthases and also present as a domain in the pentafunctional AROM polypeptide Swiss:P07547 [2]. 3-dehydroquinate (DHQ) synthase catalyses the formation of dehydroquinate (DHQ) and orthophosphate from 3-deoxy-D-arabino heptulosonic 7 phosphate [1]. This reaction is part of the shikimate pathway which is involved in the biosynthesis of aromatic amino acids.

10

Number of members: 25

933. Dihydrofolate reductase signature (DiHfolate_red)

15

Dihydrofolate reductases (EC 1.5.1.3) [1] are ubiquitous enzymes which catalyze the reduction of folic acid into tetrahydrofolic acid. They can be inhibited by a number of antagonists such as trimethoprim and methotrexate which are used as antibacterial or anticancerous agents. A signature pattern was derived from a region in the N-terminal part of these enzymes, which includes a conserved Pro-Trp dipeptide; the tryptophan has been shown [2] to be involved in the binding of substrate by the enzyme.

20

Consensus pattern[LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ] Sequences known to belong to this class detected by the patternALL, except for type II bacterial, plasmid-encoded, dihydrofolate reductases which do not belong to the same class of enzymes.

25

[1] Harpers' Review of Biochemistry, Lange, Los Altos (1985).

[2] Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C., Kraut J. J. *Biol. Chem.* 257:13650-13662(1982).

30

934. (DIL)

[1] Ponting CP; Medline: 95397417 "AF-6/cno: neither a kinesin nor a myosin, but a bit of both." Trends Biochem Sci 1995;20:265-266.

Number of members: 31

935. (DNA_gyraseB_C)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaebacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaebacteria the enzyme, known as DNA gyrase, consists of two subunits (genes *gyrA* and *gyrB* [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes *parC* and *parE*). In eukaryotes, type II topoisomerase is a homodimer.

There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

```

<-----About-1400-residues----->
[-----Protein 39-*-----][----Protein 52----] Phage T4
[-----gyrB-----*-----][-----gyrA-----] Prokaryote II
Archaeobacteria
[-----parE-----*-----][-----parD-----] Prokaryote IV
[-----*-----] Eukaryote and ASF

```

*: Position of the pattern.

As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in *gyrB*, in *parE*, and in protein 39 of phage T4 topoisomerase.

- 5 Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the pattern ALL.

[1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).

[2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).

- 10 [3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).

[4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

936. (DNA_topoisolIV)

DNA topoisomerase II signature (cross-reference = TOPOISOMERASE_II)

15 DNA topoisomerase I (EC 5.99.1.2) [1,2,3,4,E1] is one of the two types of enzyme that catalyze the interconversion of topological DNA isomers. Type II topoisomerases are ATP-dependent and act by passing a DNA segment through a transient double-strand break. Topoisomerase II is found in phages, archaeobacteria, prokaryotes, eukaryotes, and in African Swine Fever virus (ASF). In bacteriophage T4 topoisomerase II consists of three subunits (the product of genes 39, 52 and 60). In prokaryotes and in archaeobacteria the enzyme, known as DNA gyrase, consists of two subunits (genes *gyrA* and *gyrB* [E2]). In some bacteria, a second type II topoisomerase has been identified; it is known as topoisomerase IV and is required for chromosome segregation, it also consists of two subunits (genes *parC* and *parE*).
20 In eukaryotes, type II topoisomerase is a homodimer.

25 There are many regions of sequence homology between the different subtypes of topoisomerase II. The relation between the different subunits is shown in the following representation:

30 <-----About-1400-residues----->

[-----Protein 39-*-----][---Protein 52---] Phage T4

[-----gyrB-----*-----][-----gyrA-----] Prokaryote II Archaeobacteria

761

[-----parE-----*-----][-----parD-----] Prokaryote IV
 [-----*-----] Eukaryote and ASF

['*': Position of the pattern.

- 5 As a signature pattern for this family of proteins, a region was selected that contains a highly conserved pentapeptide. The pattern is located in *gyrB*, in *parE*, and in protein 39 of phage T4 topoisomerase.

10 Consensus pattern [LIVMA]-x-E-G-[DN]-S-A-x-[STAG] Sequences known to belong to this class detected by the patternALL.

- [1] Sternglanz R. Curr. Opin. Cell Biol. 1:533-535(1990).
 [2] Bjornsti M.-A. Curr. Opin. Struct. Biol. 1:99-103(1991).
 [3] Sharma A., Mondragon A. Curr. Opin. Struct. Biol. 5:39-47(1995).
 15 [4] Roca J. Trends Biochem. Sci. 20:156-160(1995).

937. Prolyl oligopeptidase family serine active site (DPPIV_N_term)

20 The prolyl oligopeptidase family [1,2,3] consist of a number of evolutionary related peptidases whose catalytic activity seems to be provided by a charge relay system similar to that of the trypsin family of serine proteases, but which evolved by independent convergent evolution. The known members of this family are listed below.

25 - Prolyl endopeptidase (EC 3.4.21.26) (PE) (also called post-proline cleaving enzyme). PE is an enzyme that cleaves peptide bonds on the C-terminal side of prolyl residues. The sequence of PE has been obtained from a mammalian species (pig) and from bacteria (*Flavobacterium meningosepticum* and *Aeromonas hydrophila*); there is a high degree of sequence conservation between these sequences.

- *Escherichia coli* protease II (EC 3.4.21.83) (oligopeptidase B) (gene *prtB*) which cleaves peptide bonds on the C-terminal side of lysyl and arginyl residues.

30 - Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV). DPP IV is an enzyme that removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline.

762

- Yeast vacuolar dipeptidyl aminopeptidase A (DPAP A) (gene: STE13) which is responsible for the proteolytic maturation of the alpha-factor precursor.

- Yeast vacuolar dipeptidyl aminopeptidase B (DPAP B) (gene: DAP2).

- 5 - Acylamino-acid-releasing enzyme (EC 3.4.19.1) (acyl-peptide hydrolase). This enzyme catalyzes the hydrolysis of the amino-terminal peptide bond of an N-acetylated protein to generate a N-acetylated amino acid and a protein with a free amino-terminus.

10 A conserved serine residue has experimentally been shown (in E.coli protease II as well as in pig and bacterial PE) to be necessary for the catalytic mechanism. This serine, which is part of the catalytic triad (Ser, His, Asp), is generally located about 150 residues away from the C-terminal extremity of these enzymes (which are all proteins that contains about 700 to 800 amino acids).

15 Consensus pattern D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2) [S is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for yeast DPAP A.

20 Note: these proteins belong to families S9A/S9B/S9C in the classification of peptidases [4,E1].

[1] Rawlings N.D., Polgar L., Barrett A.J. Biochem. J. 279:907-911(1991).

[2] Barrett A.J., Rawlings N.D. Biol. Chem. Hoppe-Seyler 373:353-360(1992).

[3] Polgar L., Szabo E. Biol. Chem. Hoppe-Seyler 373:361-366(1992).

[4] Rawlings N.D., Barrett A.J. Meth. Enzymol. 244:19-61(1994).

25 938. Deoxyhypusine synthase (DS)

30 Eukaryotic initiation factor 5A (eIF-5A) contains an unusual amino acid, hypusine [N epsilon-(4-aminobutyl-2-hydroxy)lysine]. The first step in the post-translational formation of hypusine is catalysed by the enzyme deoxyhypusine synthase (DS) EC:1.1.1.249. The modified version of eIF-5A, and DS, are required for eukaryotic cell proliferation [1].

Number of members: 9

[1] Liao DI, Wolff EC, Park MH, Davies DR; Medline: 98154315 "Crystal structure of the NAD complex of human deoxyhypusine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site." Structure 1998;6:23-32.

5

939. (DUF21)

10 Many of the sequences in this family are annotated as hemolysins, however this is due to a similarity to Swiss:Q54318 that does not contain this domain. This domain is found in the N-terminus of the proteins adjacent to two intracellular CBS domains CBS.

Number of members: 42

15 940. (DUF59)

This family includes prokaryotic proteins of unknown function. The family also includes PhaH Swiss:O84984 from *Pseudomonas putida*. PhaH forms a complex with PhaF Swiss:O84982, PhaG Swiss:O84983 and PhaI Swiss:O84985, which hydroxylates phenylacetic acid to 2-hydroxyphenylacetic acid [1]. So members of this family may all be components of ring hydroxylating complexes.

Number of members: 15

25 [1] Olivera ER, Minambres B, Garcia B, Muniz C, Moreno MA, Ferrandez A, Diaz E, Garcia JL, Luengo JM; Medline: 98263372 "Molecular characterization of the phenylacetic acid catabolic pathway in *Pseudomonas putida* U: the phenylacetyl-CoA catabolon." Proc Natl Acad Sci U S A 1998;95:6419-6424.

941. (DUF82)

30

The protein contains four conserved cysteines that may be involved in metal binding or disulphide bridges.

Number of members: 4

942. Riboflavin kinase / FAD synthetase (FAD_Synth)

This family consists part of the bifunctional enzyme riboflavin kinase / FAD synthetase.

- 5 These enzymes have both ATP:riboflavin 5'-phospho transferase and ATP:FMN-adenylyltransferase activities [1]. They catalyse the 5'-phosphorylation of riboflavin to FMN and the adenylylation of FMN to FAD [1].

CAUTION: It is not clear if this region of the enzymes catalyses either or both of the enzymatic reactions.

10 Number of members: 27

[1] Manstein DJ, Pai EF; Medline: 87057286 "Purification and characterization of FAD synthetase from *Brevibacterium ammoniagenes*." J Biol Chem 1986;261:16169-16173.

15 943. [2Fe-2S] binding domain (fer2_2)

[1] Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, Engh R, Schneider M, Hof P, Huber R; Medline: 96072968 "Crystal structure of the xanthine oxidase-related aldehyde oxidoreductase from *D. gigas*." Science 1995;270:1170-1176.

20 Number of members: 53

944. Filovirus glycoprotein (Filo_glycop)

- 25 This family includes an extracellular region from the envelope glycoprotein of Ebola and Marburg viruses. This region is also produced as a separate transcript that gives rise to a non-structural, secreted glycoprotein, which is produced in large amounts and has an unknown function [1]. Processing of this protein may be involved in viral pathogenicity [2].

Number of members: 23

- 30 [1] Volchkov VE, Feldmann H, Volchkova VA, Klenk HD; Medline: 98245155 "Processing of the Ebola virus glycoprotein by the proprotein convertase furin." Proc Natl Acad Sci U S A 1998;95:5762-5767.

[2] Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST; Medline: 96195018 "The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing." *Proc Natl Acad Sci U S A* 1996;93:3602-3607.

5 945. Frataxin-like domain (Frataxin_Cyay)

This family contains proteins that have a domain related to the globular C-terminus of Frataxin the protein that is mutated in Friedreich's ataxia. This domain is found in a family of bacterial proteins. The function of this domain is currently unknown.

10 Number of members: 12

[1] Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P; Medline: 97084946 "Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction." *Trends Neurosci* 1996;19:465-468.

15 946. (GAF)

Domain present in phytochromes and cGMP-specific phosphodiesterases.

Number of members: 296

[1] Aravind L, Ponting CP; Medline: 98094688 "The GAF domain: an evolutionary link between diverse phototransducing proteins." *Trends Biochem Sci* 1997;22:458-459.

20 947. Galaptin signature (Gal-bind_lectin)

25 All vertebrates synthesize soluble galactoside-binding lectins [1,2,3] (also known as galectins, galaptins or S-lectin). These carbohydrate-binding proteins are developmentally regulated. Although their exact physiological role is not yet clear they seem to be involved in differentiation, cellular regulation and tissue construction. The sequence of galactoside-binding lectins from electric eel (electrolectin), conger eel (congerin), chicken and a number
30 of mammalian species is known. These lectins are proteins of about 130 to 140 amino acid residues (14 Kd to 16 Kd).

A number of other proteins are known to belong to this family:

- Galectin-3 (also known as MAC-2 antigen; CBP-35 or IgE-binding protein), a 35 Kd lectin which binds immunoglobulin E and which is composed of two domains: a N-terminal domain that consist of tandem repeats of a glycine/ proline-rich sequence and a C-terminal galaptin domain.

- Galectin-4 [4], which is composed of two galaptin domains.

- Galectin-5.

- Galectin-7 [5], a keratinocyte protein which could be involved in cell-cell and/or cell-matrix interactions necessary for normal growth control.

- Galectin-8 [6], which is composed of two galaptin domains.

- Galectin-9 [7], which is composed of two galaptin domains.

- Human eosinophil lysophospholipase (EC 3.1.1.5) [8] (Charcot-Leyden crystal protein), a protein that may have both an enzymatic and a lectin activities. It forms hexagonal bipyramidal crystals in tissues and secretions from sites of eosinophil-associated inflammation.

- *Caenorhabditis elegans* 32 Kd lactose-binding lectin [9]. This lectin is composed of two galaptin domains.

- *Caenorhabditis elegans* lec-7 and lec-8.

One of the conserved regions of these lectins contains a tryptophan that has been shown [10] to be essential to the binding of galactosides. This region was used as a signature pattern for these proteins.

Consensus pattern W-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQE GSKV]-x-[GH]-x(3)-[DENKHS]-[LIVMFC] [W binds carbohydrate] Sequences known to belong to this class detected by the pattern ALL, except for pig galectin 4.

[1] Barondes S.H., Gitt M.A., Leffler H., Cooper D.N.W. *Biochimie* 70:1627-1632(1988).

[2] Hirabayashi J., Kasai K.-I. *J. Biochem.* 104:1-4(1988).

[3] Barondes S.H., Castronovo V., Cooper D.N.W., Cummings R.D., Drickamer K., Feizi T., Gitt M.A., Hirabayashi J., Hughes C., Kasai K.-I., Leffler H., Liu F.-T., Lotan R., Mercurio A.M., Monsigny M., Pillair S., Poirer F., Raz A., Rigby P.W.J., Rini J.M., Wang J.L. *Cell* 76:597-598(1994).

[4] Oda Y., Herrmann J., Gitt M., Turck C.W., Burlingame A.L., Barondes S.H., Leffler H. J. Biol. Chem. 268:5929-5939(1993).

[5] Madsen P., Rasmussen H.H., Flint T., Gromov P., Kruse T.A., Honore B., Vorum H., Celis J.E. J. Biol. Chem. 270:5823-5829(1995).

5 [6] Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y. J. Biol. Chem. 270:3447-3453(1995).

[7] Wada J., Kanwar Y.S. J. Biol. Chem. 272:6078-6086(1997).

[8] Ackerman S.J., Corrette S.E., Rosenberg H.F., Bennett J.C., Mastrianni D.M., Nicholson-Weller A., Weller P.F., Chin D.T., Tenen D.G. J. Immunol. 150:456-468(1993).

10 [9] Hirabayashi J., Satoh M., Kasai K.-I. J. Biol. Chem. 267:15485-15490(1992).

[10] Abbott W.M., Feizi T. J. Biol. Chem. 266:5552-5557(1991).

948. (GARS) Phosphoribosylglycinamide synthetase signature (phosphoribosylamine glycine ligase)

15 PROSITE: PDOC00164; cross-reference(s): PS00184

[1] catalyzes the second step in the de novo biosynthesis of purine, the ATP-dependent addition of 5-phosphoribosylamine to glycine to form 5-phosphoribosylglycinamide.

In bacteria GARS is a monofunctional enzyme (encoded by the purD gene), in of a
20 bifunctional enzyme (encoded by the ADE5,7 gene), in higher eukaryotes it is part, with AIRS and with phosphoribosylglycinamide formyltransferase (GART) of a trifunctional enzyme (GARS-AIRS-GART).

The sequence of GARS is well conserved. A highly conserved octapeptide was selected as a signature pattern.

25

Consensus pattern R-F-G-D-P-E-x-[QM]

Sequences known to belong to this class detected by the pattern ALL.

[1] Aiba A., Mizobuchi K. J. Biol. Chem. 264:21239-21246(1989).

30

949. GLTT - GLTT repeat (12 copies)

This short repeat of unknown function is found in multiple copies in several *C. elegans* proteins. The repeat is five residues long and consists of XGLTT where X can be any amino acid. Number of members: 34.

5 950. Glu_synthase - Conserved region in glutamate synthase

This family represents a region of the glutamate synthase protein. This region is expressed as a separate subunit in the glutamate synthase alpha subunit from archaeobacteria, or part of a large multidomain enzyme in other organisms. The aligned region of these proteins contains a putative FMN binding site and Fe-S cluster. Number of members: 44.

10

[1] Medline: 97082505. Sequence of the GLT1 gene from *Saccharomyces cerevisiae* reveals the domain structure of yeast glutamate synthase. Filetici P, Martegani MP, Valenzuela L, Gonzalez A, Ballario P; *Yeast* 1996;12:1359-1366.

15 951. (Glyco_hydro_2) Glycosyl hydrolases family 2 signatures

GLYCOSYL_HYDROL_F2_1; PS00608; GLYCOSYL_HYDROL_F2_2

It has been shown [1,2,E1] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

-Beta-galactosidases (EC 3.2.1.23) from bacteria such as *Escherichia coli* (genes *lacZ* and *ebgA*), *Clostridium acetobutylicum*, *Clostridium thermosulfurogenes*, *Klebsiella pneumoniae*, *Lactobacillus delbrueckii*, or *Streptococcus thermophilus* and from the fungi *Kluyveromyces fragilis*.

-Beta-glucuronidase (EC 3.2.1.31) from *Escherichia coli* (gene *uidA*) and from mammals.

One of the conserved regions in these enzymes is centered on a conserved glutamic acid residue which has been shown [3], in *Escherichia coli lacZ*, to be the general acid/base catalyst in the active site of the enzyme. This region has been used as a signature pattern. A highly conserved region located some sixty residues upstream from the active site glutamate has been selected as a second signature pattern.

30 Consensus pattern N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYWS](2)-x(3)-[DN]-x(2)-G-[LIVMFYW](4) Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [DENQLF]-[KRVW]-N-[HRY]-[STAPPV]-[SAC]-[LIVMFS](3)-W-[GS]-x(2,3)-N-E [E is the active site residue] Sequences known to belong to this class detected by the pattern ALL, except for *Rhizobium meliloti* lacZ.

- 5 [1]Henrissat B. Biochem. J. 280:309-316(1991).
- [2]Schroeder C.J., Robert C., Lenzen G., McKay L.L., Mercenier A. J. Gen. Microbiol. 137:369-380(1991).
- [3]Gebler J.C., Aebersold R., Withers S.G. J. Biol. Chem. 267:11126-11130(1992).

- 10 952. (Glyco_hydro_3) Glycosyl hydrolases family 3 active site
PROSITE: PDOC00621. PROSITE cross-reference(s)PS00775; GLYCOSYL_HYDROL_F3

It has been shown [1,2] that the following glycosyl hydrolases can be, on the basis of sequence similarities, classified into a single family:

- 15 -Beta glucosidases (EC 3.2.1.21) from the fungi *Aspergillus wentii* (A-3), *Hansenula anomala*, *Kluyveromyces fragilis*, *Saccharomycopsis fibuligera*, (BGL1 and BGL2), *Schizophyllum commune* and *Trichoderma reesei* (BGL1).
- Beta glucosidases from the bacteria *Agrobacterium tumefaciens* (Cbg1), *Butyrivibrio fibrisolvens* (bglA), *Clostridium thermocellum* (bglB), *Escherichia coli* (bglX), *Erwinia chrysanthemi* (bgxA) and *Ruminococcus albus*.
- 20 -*Alteromonas* strain O-7 beta-hexosaminidase A (EC 3.2.1.52).
- Bacillus subtilis* hypothetical protein yzbA.
- Escherichia coli* hypothetical protein ycfO and HI0959, the corresponding *Haemophilus influenzae* protein.

- 25 One of the conserved regions in these enzymes is centered on a conserved aspartic acid residue which has been shown [3], in *Aspergillus wentii* beta-glucosidase A3, to be implicated in the catalytic mechanism. This region was used as a signature pattern.

Consensus pattern[LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-[SGADNI] [D is the active site residue]

- 30 Sequences known to belong to this class detected by the patternALL.

- [1]Henrissat B. Biochem. J. 280:309-316(1991).
- [2]Castle L.A., Smith K.D., Morris R.O. J. Bacteriol. 174:1478-1486(1992).

[3]Bause E., Legler G. *Biochim. Biophys. Acta* 626:459-465(1980).

953. GP120 - Envelope glycoprotein GP120

The entry of HIV requires interaction of viral GP120 with Swiss:P01730 and a chemokine receptor on the cell surface. Number of members: 17891

[1]Medline: 98303379. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA; *Nature* 1998;393:648-659.

954. (GSP11_E) Bacterial type II secretion system protein E signature

PROSITE: PDOC00567. PROSITE cross-reference(s) PS00662; T2SP_E

A number of bacterial proteins, some of which are involved in a general secretion pathway (GSP) for the export of proteins (also called the type II pathway) [1,2], have been found to be evolutionary related. These proteins are listed below:

-The 'E' protein from the GSP operon of: *Aeromonas* (gene *exeE*); *Erwinia* (gene *outE*); *Escherichia coli* (gene *yeH*); *Klebsiella pneumoniae* (gene *puE*); *Pseudomonas aeruginosa* (gene *xcpR*); *Vibrio cholerae* (gene *epsE*) and *Xanthomonas campestris* (gene *xpsE*).

-*Agrobacterium tumefaciens* Ti plasmid *virB* operon protein 11. This protein is required for the transfer of T-DNA to plants.

-*Bacillus subtilis* *comG* operon protein 1 which is required for the uptake of DNA by competent *Bacillus subtilis* cells.

-*Aeromonas hydrophila* *tapB*, involved in type IV pilus assembly.

-*Pseudomonas* protein *pilB*, which is essential for the formation of the pili.

-*Pseudomonas aeruginosa* protein twitching mobility protein *pilT*.

-*Neisseria gonorrhoeae* type IV pilus assembly protein *pilF*.

-*Vibrio cholerae* protein *tcpT*, which is involved in the biosynthesis of the *tcp* pilus.

-*Escherichia coli* protein *hopB* (*hopB*).

-*Escherichia coli* hypothetical protein *ycgB*.

-*Escherichia coli* hypothetical protein *yggR*.

These proteins have from 344 (*pilT* and *virB11*) to 568 (*tapB*) amino acids, they are probably cytoplasmically located and, on the basis of the presence of a conserved P-loop

771

region (see <PDOC00017>), probably bind ATP. A region that overlaps the 'B' motif of ATP-binding proteins was selected as a signature pattern.

Consensus pattern[LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D

5 Sequences known to belong to this class detected by the patternALL, except for ygcB.

[1]Salmond G.P.C., Reeves P.J. Trends Biochem. Sci. 18:7-12(1993).

[2]Hobbs M., Mattick J.S. Mol. Microbiol. 10:233-243(1993).

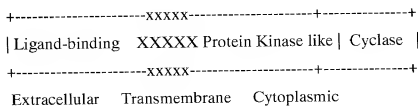
10 955. (guanylate_cyc) Guanylate cyclases signature

PROSITE: PDOC00425. PROSITE cross-reference(s) PS00452;

GUANYLATE_CYCLASES Guanylate cyclases (EC 4.6.1.2) [1 to 4] catalyze the formation of cyclic GMP (cGMP) from GTP. cGMP acts as an intracellular messenger, activating cGMP dependent kinases and regulating CGMP-sensitive ion channels. The role of cGMP as a second messenger in vascular smooth muscle relaxation and retinal photo-transduction is well established. Guanylate cyclase is found both in the soluble and particular fraction of eukaryotic cells. The soluble and plasma membrane-bound forms differ in structure, regulation and other properties.

20 Most currently known plasma membrane-bound forms are receptors for small polypeptides. The topology of such proteins is the following: they have a N-terminal extracellular domain which acts as the ligand binding region, then a transmembrane domain, followed by a large cytoplasmic C-terminal region that can be subdivided into two domains: a protein kinase-like domain that appears important for proper signalling and a cyclase catalytic domain. This topology is schematically represented below.

25



30

The known guanylate cyclase receptors are:

-The sea-urchins receptors for speract and resact, which are small peptides that stimulate sperm motility and metabolism.

-The receptors for natriuretic peptides (ANF). Two forms of ANF receptors with guanylate cyclase activity are currently known: GC-A (or ANP-A) which seems specific to atrial natriuretic peptide (ANP), and GC-B (or ANP-B) which seems to be stimulated more effectively by brain natriuretic peptide (BNP) than by ANP.

-The receptor for Escherichia coli heat-stable enterotoxin (GC-C). The endogenous ligand for this intestinal receptor seems to be a small peptide called guanylin.

-Retinal guanylate cyclase (retGC) which probably plays a specific functional role in the rods and/or cones of photoreceptors. It is not known if this protein acts as receptor, but its structure is similar to that of the other plasma membrane-bound GCs.

The soluble forms of guanylate cyclase are cytoplasmic heterodimers. The two subunits, alpha and beta are proteins of from 70 to 82 Kd which are highly related. Two forms of beta subunits are currently known: beta-1 which seems to be expressed in lung and brain, and beta-2 which is more abundant in kidney and liver.

The membrane and cytoplasmic forms of guanylate cyclase share a conserved domain which is probably important for the catalytic activity of the enzyme. Such a domain is also found twice in the different forms of membrane-bound adenylate cyclases (also known as class-III) [5,6] from mammals, slime mold or Drosophila. A consensus pattern was derived from the most conserved region in that domain.

Consensus pattern G-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-[DNTA]-x(5)-[DE]

Sequences known to belong to this class detected by the pattern ALL, except for the sea urchin Arbacia punctulata resact receptor which lack this domain.

Note this pattern will detect both domains of adenylate cyclases class-III.

[1]Koesling D., Boehme E., Schultz G. FASEB J. 5:2785-2791(1991).

[2]Garbers D.L. New Biol. 2:499-504(1990).

[3]Garbers D.L. Cell 71:1-4(1992).

[4]Yuen P.S.T., Garbers D.L. Annu. Rev. Neurosci. 15:193-225(1992).

[5]Iyengar R. FASEB J. 7:768-775(1993).

[6]Barzu O., Danchin A. Prog. Nucleic Acid Res. Mol. Biol. 49:241-283(1994).

956. Hemolysin-type calcium-binding region signature (HemolysinCabinD)

Gram-negative bacteria produce a number of proteins which are secreted into the growth medium by a mechanism that does not require a cleaved N-terminal signal sequence. These proteins, while having different functions, seem [1] to share two properties: they bind calcium and they contain a variable number of tandem repeats consisting of a nine amino acid motif rich in glycine, aspartic acid and asparagine. It has been shown [2] that such a domain is involved in the binding of calcium ions in a parallel beta roll structure. The proteins which are currently known to belong to this category are:

- Hemolysins from various species of bacteria. Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The hemolysins which are known to contain such a domain are those from: *E. coli* (gene hlyA), *A. pleuropneumoniae* (gene appA), *A. actinomycetemcomitans* and *P. haemolytica* (leukotoxin) (gene lktA).

- Cyclolysin from *Bordetella pertussis* (gene cyaA). A multifunctional protein which is both an adenylate cyclase and a hemolysin.

- Extracellular zinc proteases: serralyisin (EC 3.4.24.40) from *Serratia*, prtB and prtC from *Erwinia chrysanthemi* and aprA from *Pseudomonas aeruginosa*.

- Nodulation protein nodO from *Rhizobium leguminosarum*.

A signature pattern was derived from conserved positions in the sequence of the calcium-binding domain.

Consensus pattern D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D Sequences known to belong to this class detected by the pattern ALL.

Note: This pattern is found once in nodO and the extracellular proteases but up to 5 times in some hemolysin/cyclolysins.

[1] Economou A., Hamilton W.D.O., Johnston A.W.B., Downie J.A. EMBO J. 9:349-354(1990).

[2] Baumann U., Wu S., Flaherty K.M., McKay D.B. EMBO J. 12:3357-3364(1993).

957. Hint module (Hint)

This is an alignment of the Hint module in the Hedgehog proteins. It does not include any Inteins which also possess the Hint module.

Number of members: 36

- 5 [1] Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ; Medline: 97474313
 “Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins.” Cell 1997;91:85-97.

958. Hydantoinase/oxoprolinase (Hydantoinase)

10

This family includes the enzymes hydantoinase and oxoprolinase EC:3.5.2.9. Both reactions involve the hydrolysis of 5-membered rings via hydrolysis of their internal imide bonds [1].

Number of members: 14

- 15 [1] Ye GJ, Breslow EB, Meister A, Guo-jie GE\$[corrected to Ye GJ]; Medline: 97113037
 “The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA cloning”
 [published erratum appears in J Biol Chem 1997 Feb 14;272(7):4646] J Biol Chem
 1996;271:32293-32300.

20 959. IMP dehydrogenase / GMP reductase signature (IMPDH_N)

IMP dehydrogenase (EC 1.1.1.205) (IMPDH) catalyzes the rate-limiting reaction of de novo GTP biosynthesis, the NAD-dependent reduction of IMP into XMP [1]. Inhibition of IMP dehydrogenase activity results in the cessation of DNA synthesis. As IMP dehydrogenase is
 25 associated with cell proliferation, it is a possible target for cancer chemotherapy. Mammalian and bacterial IMPDHs are tetramers of identical chains. There are two IMP dehydrogenase isozymes in humans [2].

- 30 GMP reductase (EC 1.6.6.8) catalyzes the irreversible and NADPH-dependent reductive deamination of GMP into IMP [3]. It converts nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and maintains intracellular balance of A and G nucleotides.

775

IMP dehydrogenase and GMP reductase share many regions of sequence similarity. One of these regions is centered on a cysteine residue thought [3] to be involved in binding IMP. This region was used as a signature pattern.

- 5 Consensus pattern[LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T [C is the putative IMP-binding residue] Sequences known to belong to this class detected by the pattern ALL.

[1] Collart F.R., Huberman E. J. Biol. Chem. 263:15769-15772(1988).

- 10 [2] Natsumeda Y., Ohno S., Kawasaki H., Konno Y., Weber G., Suzuki K. J. Biol. Chem. 265:5292-5295(1990).

[3] Andrews S.C., Guest J.R. Biochem. J. 255:35-43(1988).

960. impB/mucB/samB family (IMS)

15 These proteins are involved in UV protection (Swiss).

Number of members: 38

961. Type II intron maturase (Intron_maturas2)

20 Group II introns use intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA [2]. Although this type of intron is self splicing in vitro they require a maturase protein for splicing in vivo. It has been shown that a specific region of the aI2 intron is needed for the maturase function [1]. This region was found to be conserved in group II introns and called domain X [3].

25 Number of members: 335

[1] Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P;

- 30 Medline: 94301788 "Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. Nucleic Acids Res 1994;22:2057-2064.

[2] Guo H, Zimmerly S, Perlman PS, Lambowitz AM; Medline: 98031910 "Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA." EMBO J 1997;16:6835-6848.

[3] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 "Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

962. LAGLIDADG endonuclease (Intron_maturase)

- [1] Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL; Medline: 97331323 "The structure of I-CreI, a group I intron-encoded homing endonuclease." Nat Struct Biol 1997;4:468-476.
- [2] Belfort M, Roberts RJ; Medline: 97402526 "Homing endonucleases: keeping the house in order." Nucleic Acids Res 1997;25:3379-3388.
- [3] Dalggaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS; Medline: 98026854 "Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res 1997;25:4626-4638.

Number of members: 220

963. Isopentenyl transferase (IPT)

Isopentenyl transferase / dimethylallyl transferase synthesizes isopentenyladenosine 5'-monophosphate, a cytokinin that induces shoot formation on host plants infected with the Ti plasmid [1].

Number of members: 16

- [1] Canaday J, Gerad JC, Crouzet P, Otten L; Medline: 93101133 "Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4." Mol Gen Genet 1992;235:292-303.

964. Laminin EGF-like (Domains III and V) (laminin_EGF)

This family is like EGF but has 8 conserved cysteines instead of 6.

Number of members: 501

[1] Engel J; Medline: 93041759 "Laminins and other strange proteins." Biochemistry
1992;31:10643-10651.

965. Legume lectins signatures (lectin_legA)

Leguminous plants synthesize sugar-binding proteins which are called legume lectins [1,2].

These lectins are generally found in the seeds. The exact function of legume lectins is not known but they may be involved in the attachment of nitrogen-fixing bacteria to legumes and in the protection against pathogens. Legume lectins bind calcium and manganese (or other transition metals).

Legume lectins are synthesized as precursor proteins of about 230 to 260 amino acid residues. Some legume lectins are proteolytically processed to produce two chains: beta (which corresponds to the N-terminal) and alpha (C-terminal). The lectin concanavalin A (conA) from jack bean is exceptional in that the two chains are transposed and ligated (by formation of a new peptide bond). The N-terminus of mature conA thus corresponds to that of the alpha chain and the C-terminus to the beta chain.

Two signature patterns were developed specific to legume lectins: the first is located in the C-terminal section of the beta chain and contains a conserved aspartic acid residue important for the binding of calcium and manganese; the second one is located in the N-terminal of the alpha chain.

Consensus pattern [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST] [D binds manganese and calcium] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIVF]-G-[LF]-[ST] Sequences known to belong to this class detected by the pattern ALL.

[1] Sharon N., Lis H. FASEB J. 4:3198-320(1990).

[2] Lis H., Sharon N. Annu. Rev. Biochem. 55:33-37(1986).

966. Malate synthase signature (malate_synthase)

5 Malate synthase (EC 4.1.3.2) catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form malate - the second step of the glyoxylate bypass, an alternative to the tricarboxylic acid cycle in bacteria, fungi and plants. Malate synthase is a protein of 530 to 570 amino acids whose sequence is highly conserved across species [1]. As a signature pattern, a very conserved region was selected in the central section of the enzyme.

10

Consensus pattern[**KR**]-[**DENQ**]-H-x(2)-G-L-N-x-G-x-W-D-Y-[**LIVM**]-F Sequences known to belong to this class detected by the pattern **ALL**.

[1] Bruinenberg P.G., Blaauw M., Kazemier B., Ab G. Yeast 6:245-254(1990).

15

967. MatK/TrnK amino terminal region (MatK_N)

[1] Mohr G, Perlman PS, Lambowitz AM; Medline: 94077696 "Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function." Nucleic Acids Res 1993;21:4991-4997.

20

Number of members: 495

968. MOZ/SAS family (MOZ_SAS)

25

This region of these proteins has been suggested to be homologous to acetyltransferases [1]. However the similarity is not supported by standard sequence analysis.

Number of members: 15

30

[1] Kamine J, Elangovan B, Subramanian T, Coleman D, Chinnadurai G; Medline: 96182937 "Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator." Virology 1996;216:357-366.

[2] Reifsnyder C, Lowell J, Clarke A, Pillus L; Medline: 96376969 "Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases" [see comments] [published erratum appears in Nat Genet 1997 May;16(1):109] Nat Genet 1996;14:42-49.

969. mRNA capping enzyme (mRNA_cap_enzyme)

[1] Hakansson K, Doherty AJ, Shuman S, Wigley DB; Medline: 97304383 "X-ray crystallography reveals a large conformational change during guanylyl transfer by mRNA capping enzymes." Cell 1997;89:545-553.

Number of members: 7

970. DNA mismatch repair proteins mutS family signature (MutS_C)

Mismatch repair contributes to the overall fidelity of DNA replication [1]. It involves the correction of mismatched base pairs that have been missed by the proofreading element of the DNA polymerase complex. The sequence of some proteins involved in mismatch repair in different organisms have been found to be evolutionary related [2,3]. One of these families is called mutS [4,E1], it consists of:

- Prokaryotic protein mutS protein (also called hexA in *Streptococcus pneumoniae*). Muts is thought to carry out the mismatch recognition step of DNA repair.
- Eukaryotic MSH1, which is involved in mitochondrial DNA repair.
- Eukaryotic MSH2, which is involved in nuclear postreplication mismatch repair. MSH2 heterodimerizes with MSH6. In man, MSH2 is involved in a form of familial hereditary nonpolyposis colon cancer (HNPCC).
- Eukaryotic MSH3, which is probably involved in the repair of large loops.
- Eukaryotic MSH4, which is involved in meiotic recombination.
- Eukaryotic MSH5, which is involved in meiotic recombination.
- Eukaryotic MSH6 (also known as G/T mismatch binding protein), a DNA-repair protein that binds to G/T mismatches through heterodimerization with MSH2.
- Prokaryotic protein mutS2 whose function is not yet known.
- A coral (*Sarcophyton glaucum*) mitochondrial encoded mutS-like protein.

As a signature pattern for this class of mismatch repair proteins a region rich in glycine and negatively charged residues was selected. This region is found in the C-terminal section of these proteins; about 80 residues to the C-terminal of an ATP-binding site motif 'A' (P-loop) (see <PDOC00017>).

Consensus pattern[ST]-[LIVMF]-x-[LIVM]-x-D-E-[LIVMFY]-[GC]-[RKH]-G-[GST]-x(4)-G Sequences known to belong to this class detected by the pattern ALL, except for mutS2.

[1] Modrich P. Annu. Rev. Biochem. 56:435-466(1987).

[2] Haber L.T., Walker G.C. EMBO J. 10:2707-2715(1991).

[3] New L., Liu K., Crouse G.F. Mol. Gen. Genet. 239:97-108(1993).

[4] Eisen J.A. Nucleic Acids Res. 26:4291-4300(1998).

971. MutS family, N-terminal putative DNA binding domain (MutS_N)

This family consists of the N-terminal region of proteins in the mutS family of DNA mismatch repair proteins and is found associated with MutS_C located in the C-terminal region. The mutS family of proteins is named after the salmonella typhimurium MutS protein involved in mismatch repair; other members of the family included the eukaryotic MSH 1,2,3,4,5 and 6 proteins. These have various roles in DNA repair and recombination. Human MSH has been implicated in non-polyposis colorectal carcinoma (HNPCC) and is a mismatch binding protein [2]. The aligned region corresponds in part with domains A1, A2 (which may bind DNA) and B (which binds dsDNA in vitro) from T. thermophilus MutS as characterised in [1].

Number of members: 43

972. Domain in Myosin and Kinesin Tails (MyTH4)

Domain present twice in myosin-VIIIa, and also present in 3 other myosins.

[1] Chen ZY, Hasson T, Kelley PM, Schwender BJ, Schwartz MF, Ramakrishnan M, Kimberling WJ, Mooseker MS, Corey DP; Medline: 97038686 "Molecular cloning and

domain structure of human myosin-VIIa, the gene product defective in Usher syndrome 1B." Genomics 1996;36:440-448.

Number of members: 21

973. Sodium and potassium ATPases beta subunits signatures (Na_K-ATPase)

The sodium pump (Na⁺,K⁺ ATPase), located in the plasma membrane of all animal cells [1], is an heterotrimer of a catalytic subunit (alpha chain), a glycoprotein subunit of about 34 Kd (beta chain) and a small hydrophobic protein of about 6 Kd. The beta subunit seems [2] to regulate, through the assembly of alpha/beta heterodimers, the number of sodium pumps transported to the plasma membrane.

Structurally the beta subunit is composed of a charged cytoplasmic domain of about 35 residues, followed by a transmembrane region, and a large extracellular domain that contains three disulfide bonds and glycosylation sites. This structure is schematically represented in the figure below.

+----+ +-+ +-----+ |||||

xxxxxxxxxxxxxxxxxxxxxxxxCxxxxCxCxxCxxxxxxxxCxxxxxxxxCxxxx

*** <-Cyt-><TM><-----Extracellular----->

'C': conserved cysteine involved in a disulfide bond.

'*': position of the patterns.

Two isoforms of the beta subunit (beta-1 and beta-2) are currently known; they share about 50% sequence identity. Gastric (K⁺, H⁺) ATPase (proton pump) responsible for acid production in the stomach consist of two subunits [3]; the beta chain is highly similar to the sodium pump beta subunits. Two signature patterns were developed for beta subunits. The first is located in the cytoplasmic domain, while the second is found in the extracellular domain and contains two of the cysteines involved in disulfide bonds.

Consensus pattern [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W

Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G [The two C's are involved in disulfide bonds] Sequences known to belong to this class detected by the patternALL, except for the beta subunit of the sodium pump of brine shrimp whose sequence is highly divergent in that region.

[1] Horisberger J.D., Lemas V., Krahenbul J.P., Rossier B.C. Annu. Rev. Physiol. 53:565-584(1991).

[2] McDonough A.A., Gerring K., Farley R.A. FASEB J. 4:1598-1605(1990).

[3] Toh B.-H., Gleeson P.A., Simpson R.J., Moritz R.L., Callaghan J.M., Goldkorn I., Jones C.M., Martinelli T.M., Mu F.-T., Humphris D.C., Pettitt J.M., Mori Y., Masuda T., Sobieszczuk P., Weinstock J., Mantamadiotis T., Baldwin G.S. Proc. Natl. Acad. Sci. U.S.A. 87:6418-6422(1990).

974. Respiratory-chain NADH dehydrogenase subunit 1 signatures (NADHdh)

Respiratory-chain NADH dehydrogenase (EC 1.6.5.3) [1,2] (also known as complex I or NADH-ubiquinone oxidoreductase) is an oligomeric enzymatic complex located in the inner mitochondrial membrane which also seems to exist in the chloroplast and in cyanobacteria (as a NADH-plastoquinone oxidoreductase). Among the 25 to 30 polypeptide subunits of this bioenergetic enzyme complex there are fifteen which are located in the membrane part, seven of which are encoded by the mitochondrial and chloroplast genomes of most species. The most conserved of these organelle-encoded subunits is known as subunit 1 (gene ND1 in mitochondrion, and NDH1 in chloroplast) and seems to contain the ubiquinone binding site.

The ND1 subunit is highly similar to subunit 4 of *Escherichia coli* formate hydrogenlyase (gene hycD), subunit C of hydrogenase-4 (gene hycC). *Paracoccus denitrificans* NQO8 and *Escherichia coli* nuoH NADH-ubiquinone oxidoreductase subunits also belong to this family [3]. Two signature patterns were developed based on conserved regions of this subunit.

Consensus pattern G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]-[LIVMFTA]-K-[LVMYST]-[LIVMFYG]-x-[KR]-[EQG] Sequences known to belong to this class detected by the patternALL, except for watermelon and *Leishmania* ND1.

Consensus pattern P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G
Sequences known to belong to this class detected by the pattern ALL, except for
Chlamydomonas reinhardtii and *Pisaster ochraceus* ND1, and tobacco NDH1.

[1] Ragan C.I. Curr. Top. Bioenerg. 15:1-36(1987).

[2] Weiss H., Friedrich T., Hofhaus G., Preis D. Eur. J. Biochem. 197:563-576(1991).

[3] Weidner U., Geier S., Ptocek A., Friedrich T., Leif H., Weiss H. J. Mol. Biol. 233:109-122(1993).

975. Nickel-dependent hydrogenases large subunit signatures (NiFeSe_Hases)

Hydrogenases are enzymes that catalyze the reversible activation of hydrogen and which occur widely in prokaryotes as well as in some eukaryotes. There are various types of hydrogenases, but all of them seem to contain at least one iron-sulfur cluster. They can be broadly divided into two groups: hydrogenases containing nickel and, in some cases, also selenium (the [NiFe] and [NiFeSe] hydrogenases) and those lacking nickel (the [Fe] hydrogenases).

The [NiFe] and [NiFeSe] hydrogenases are heterodimer that consist of a small subunit that contains a signal peptide and a large subunit. All the known large subunits seem to be evolutionary related [1]; they contain two Cys-x-x- Cys motifs; one at their N-terminal end; the other at their C-terminal end. These four cysteines are involved in the binding of nickel [2]. In the [NiFeSe] hydrogenases the first cysteine of the C-terminal motif is a selenocysteine which has experimentally been shown to be a nickel ligand [3]. Two patterns were developed which are centered on the Cys-x-x-Cys motifs.

Alcaligenes eutrophus possess a NAD-reducing cytoplasmic hydrogenase (hoxS) [4]; this enzyme is composed of four subunits. Two of these subunits (beta and delta) are responsible for the hydrogenase reaction and are evolutionary related to the large and small subunits of membrane-bound hydrogenases. The alpha subunit of coenzyme F420 hydrogenase (EC 1.12.99.1) (FRH) from archaeobacterial methanogens also belongs to this family.

Consensus pattern R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C [The two C's are nickel ligands] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H [The two C's are nickel ligands]

5 Sequences known to belong to this class detected by the pattern ALL.

[1] Menon N.K., Robbins J., Peck H.D. Jr., Chatelus C.Y., Choi E.-S., Przybyla A.E. J. Bacteriol. 172:1969-1977(1990).

[2] Volbeda A., Charon M.-H., Piras C., Hatchikian E.C., Frey M., Fontecilla-Camps J.C.
10 Nature 373:580-587(1995).

[3] Eidsness M.K., Scott R.A., Prickrill B., der Vartanian D.V., LeGall J., Moura I., Moura J.J.G., Peck H.D. Jr. Proc. Natl. Acad. Sci. U.S.A. 86:147-151(1989).

[4] Tran-Betcke A., Warnecke U., Boecker C., Zaborosch C., Friedrich B. J. Bacteriol.
15 172:2920-2929(1990).

976. NADH-Ubiquinone oxidoreductase (complex I), chain 5 C-terminus (oxidored_q1_C)

This sub-family represents a carboxyl terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 from chloroplasts are in this family. This sub-family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.
20

Number of members: 572

[1] Walker JE; Medline: 93110040 "The NADH:ubiquinone oxidoreductase (complex I) of
25 respiratory chains." Q Rev Biophys 1992;25:253-324.

977. NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus (oxidored_q1_N)

This sub-family represents an amino terminal extension of oxidored_q1. Only NADH-Ubiquinone chain 5 and eubacterial chain L are in this family. This sub-family is part of complex I which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane.
30

Number of members: 546

[1] Walker JE; Medline: 93110040 "The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains." Q Rev Biophys 1992;25:253-324.

- 5 978. oxidored_q2. NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4L (EC 1.6.5.3). ND4L OR NAD4L. Arabidopsis thaliana (Mouse-ear cress). Mitochondrion. OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.

10

[1] SEQUENCE FROM N.A. MEDLINE; 93156682. Brandt P., Sunkel S., Unseld M., Brennicke A., Knoop V.; "The nad4L gene is encoded between exon c of nad5 and orf25 in the Arabidopsis mitochondrial genome."; Mol. Gen. Genet. 236:33-38(1992).

15

[2] SEQUENCE FROM N.A. STRAIN=CV. COLUMBIA; MEDLINE; 97141919 Unseld M., Marienfeld J.R., Brandt P., Brennicke A.; "The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides."; Nat. Genet. 15:57-61(1997).

20

979. oxidored_q4. Protein name NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 3, CHLOROPLAST. Synonym(s)EC 1.6.5.3. Gene name(s)NDHC OR NDH3 From Zea mays (Maize) Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.

SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.

25

[1] SEQUENCE FROM N.A. MEDLINE; 89281491. Steinmueller K., Ley A.C., Steinmetz A.A., Sayre R.T., Bogorad L.; "Characterization of the ndhC-psbG-ORF157/159 operon of maize plastid DNA and of the cyanobacterium Synechocystis sp. PCC6803."; Mol. Gen. Genet. 216:60-69(1989).

30

[2] SEQUENCE FROM N.A. MEDLINE; 95395841. Maier R.M., Neckermann K., Igloi G.L., Koessel H.; "Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing."; J. Mol. Biol. 251:614-628(1995).

980. PAC: PAC motif

PAC motif occurs C-terminal to a subset of all known PAS motifs. It is proposed to contribute to the PAS domain fold [3]. Number of members: 181

[1] Medline: 97446881 PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. Zhulin IB, Taylor BL, Dixon R; Trends Biochem Sci 1997;22:331-333.

[2] Medline: 95275818. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. Borgstahl GE, Williams DR, Getzoff ED; Biochemistry 1995;34:6278-6287.

[3] Medline: 98044337. PAS: a multifunctional domain family comes to light. Ponting CP, Aravind L; Curr Biol 1997;7:674-677.

981. PARP: Poly(ADP-ribose) polymerase catalytic region.

Poly(ADP-ribose) polymerase catalyses the covalent attachment of ADP-ribose units from NAD⁺ to itself and to a limited number of other DNA binding proteins, which decreases their affinity for DNA. Poly(ADP-ribose) polymerase is a regulatory component induced by DNA damage.

The carboxyl-terminal region is the most highly conserved region of the protein. Experiments have shown that a carboxyl 40 kDa fragment is still catalytically active [2]. Number of members: 19

[1] Medline: 96353841 Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE; Proc Natl Acad Sci U S A 1996;93:7481-7485.

[2] Medline: 93293867 The carboxyl-terminal domain of human poly(ADP-ribose) polymerase. Overproduction in Escherichia coli, large scale purification, and characterization. Simonin F, Hofferer L, Panzeter PL, Muller S, de Murcia G, Althaus FR; J Biol Chem 1993;268:13454-13461.

982. PC_rep: Proteasome/cyclosome repeat

[1] Medline: 97348748 A repetitive sequence in subunits of the 26S proteasome and 20S cyclosome (anaphase-promoting complex). Lupas A, Baumeister W, Hofmann K; Trends Biochem Sci 1997;22:195-196.

Number of members: 112

983. Peptidase_M1: Peptidase family M1

Members of this family are aminopeptidases. The members differ widely in specificity, hydrolysing acidic, basic or neutral N-terminal residues. This family includes leukotriene-A4 hydrolase Swiss:P09960, this enzyme also has an aminopeptidase activity [1]. Number of members: 72

[1] Medline: 95405261 Evolutionary families of metallopeptidases. Rawlings ND, Barrett AJ; Meth Enzymol 1995;248:183-228.

984. Neutral zinc metallopeptidases, zinc-binding region signature (Peptidase_M8) PROSITE cross-reference(s) PS00142; ZINC_PROTEASE

The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure [1,2,3] in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily, known as the metzincins, on the basis of this sequence similarity. They can be classified into a number of distinct families [4,E1] which are listed below along with the proteases which are currently known to belong to these families.

Family M1

- Bacterial aminopeptidase N (EC 3.4.11.2) (gene pepN).
- Mammalian aminopeptidase N (EC 3.4.11.2).
- Mammalian glutamyl aminopeptidase (EC 3.4.11.7) (aminopeptidase A). It may play a role in regulating growth and differentiation of early B-lineage cells.
- Yeast aminopeptidase yscII (gene APE2).
- Yeast alanine/arginine aminopeptidase (gene AAP1).
- Yeast hypothetical protein YIL137c.

- Leukotriene A-4 hydrolase (EC 3.3.2.6). This enzyme is responsible for the hydrolysis of an epoxide moiety of LTA-4 to form LTB-4; it has been shown that it binds zinc and is capable of peptidase activity.

Family M2

- 5 - Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE) the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which has two active centers.

Family M3

- 10 - Thimet oligopeptidase (EC 3.4.24.15), a mammalian enzyme involved in the cytoplasmic degradation of small peptides.
- Neurolysin (EC 3.4.24.16) (also known as mitochondrial oligopeptidase M or microsomal endopeptidase).
- Mitochondrial intermediate peptidase precursor (EC 3.4.24.59) (MIP). It is involved the second stage of processing of some proteins imported in the mitochondrion.
- 15 - Yeast saccharolysin (EC 3.4.24.37) (proteinase yscD).
- Escherichia coli and related bacteria dipeptidyl carboxypeptidase (EC 3.4.15.5) (gene dcp).
- Escherichia coli and related bacteria oligopeptidase A (EC 3.4.24.70) (gene opdA or prlC).
- Yeast hypothetical protein YKL134c.

20 Family M4

- Thermostable thermolysins (EC 3.4.24.27), and related thermolabile neutral proteases (bacillolysins) (EC 3.4.24.28) from various species of *Bacillus*.
- Pseudolysin (EC 3.4.24.26) from *Pseudomonas aeruginosa* (gene lasB).
- Extracellular elastase from *Staphylococcus epidermidis*.
- 25 - Extracellular protease prt1 from *Erwinia carotovora*.
- Extracellular minor protease smp from *Serratia marcescens*.
- Vibriolysin (EC 3.4.24.25) from various species of *Vibrio*.
- Protease prtA from *Listeria monocytogenes*.
- Extracellular proteinase proA from *Legionella pneumophila*.

30

Family M5

- Mycolysin (EC 3.4.24.31) from *Streptomyces cacaoi*.

Family M6

- Immune inhibitor A from *Bacillus thuringiensis* (gene ina). Ina degrades two classes of insect antibacterial proteins, attacins and cecropins.

5 Family M7

- *Streptomyces* extracellular small neutral proteases

Family M8

10 - Leishmanolysin (EC 3.4.24.36) (surface glycoprotein gp63), a cell surface protease from various species of *Leishmania*.

Family M9

- Microbial collagenase (EC 3.4.24.3) from *Clostridium perfringens* and *Vibrio alginolyticus*.

15 Family M10A

- Serralysin (EC 3.4.24.40), an extracellular metalloprotease from *Serratia*.

- Alkaline metalloproteinase from *Pseudomonas aeruginosa* (gene aprA).

- Secreted proteases A, B, C and G from *Erwinia chrysanthemi*.

20 - Yeast hypothetical protein YIL108w.

Family M10B

25 - Mammalian extracellular matrix metalloproteinases (known as matrixins) [5]: MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase).

30 - Sea urchin hatching enzyme (envelysin) (EC 3.4.24.12). A protease that allows the embryo to digest the protective envelope derived from the egg extracellular matrix.

- Soybean metalloendoproteinase 1.

Family M11

- *Chlamydomonas reinhardtii* gamete lytic enzyme (GLE).

Family M12A

- Astacin (EC 3.4.24.21), a crayfish endoprotease.
- 5 - Meprin A (EC 3.4.24.18), a mammalian kidney and intestinal brush border metalloendopeptidase.
- Bone morphogenic protein 1 (BMP-1), a protein which induces cartilage and bone formation and which expresses metalloendopeptidase activity. The *Drosophila* homolog of BMP-1 is the dorsal-ventral patterning protein tolloid.
- 10 - Blastula protease 10 (BP10) from *Paracentrotus lividus* and the related protein SpAN from *Strongylocentrotus purpuratus*.
- *Caenorhabditis elegans* protein toh-2.
- *Caenorhabditis elegans* hypothetical protein F42A10.8.
- Choriolyins L and H (EC 3.4.24.67) (also known as embryonic hatching proteins LCE and HCE) from the fish *Oryzias latipes*. These proteases participate in the breakdown of the egg envelope, which is derived from the egg extracellular matrix, at the time of hatching.
- 15

Family M12B

- 20 - Snake venom metalloproteinases [6]. This subfamily mostly groups proteases that act in hemorrhage. Examples are: adamalysin II (EC 3.4.24.46), atrolysin C/D (EC 3.4.24.42), atrolysin E (EC 3.4.24.44), fibrolase (EC 3.4.24.72), trimerelysin I (EC 3.4.25.52) and II (EC 3.4.25.53).
- Mouse cell surface antigen MS2.

25

Family M13

- Mammalian neprilysin (EC 3.4.24.11) (neutral endopeptidase) (NEP).
- Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which process the precursor of endothelin to release the active peptide.
- 30 - Kell blood group glycoprotein, a major antigenic protein of erythrocytes. The Kell protein is very probably a zinc endopeptidase.
- Peptidase O from *Lactococcus lactis* (gene pepO).

Family M27

- Clostridial neurotoxins, including tetanus toxin (TeTx) and the various botulinum toxins (BoNT). These toxins are zinc proteases that block neurotransmitter release by proteolytic cleavage of synaptic proteins such as synaptobrevins, syntaxin and SNAP-25 [7,8].

Family M30

- *Staphylococcus hyicus* neutral metalloprotease.

Family M32

- Thermostable carboxypeptidase 1 (EC 3.4.17.19) (carboxypeptidase Taq), an enzyme from *Thermus aquaticus* which is most active at high temperature.

Family M34

- Lethal factor (LF) from *Bacillus anthracis*, one of the three proteins composing the anthrax toxin.

Family M35

- Deuterolysin (EC 3.4.24.39) from *Penicillium citrinum* and related proteases from various species of *Aspergillus*.

Family M36

- Extracellular elastolytic metalloproteinases from *Aspergillus*.

From the tertiary structure of thermolysin, the position of the residues acting as zinc ligands and those involved in the catalytic activity are known. Two of the zinc ligands are histidines which are very close together in the sequence; C-terminal to the first histidine is a glutamic acid residue which acts as a nucleophile and promotes the attack of a water molecule on the carbonyl carbon of the substrate. A signature pattern which includes the two histidine and the glutamic acid residues is sufficient to detect this superfamily of proteins.

Consensus pattern[GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-
[LIVMFYWGSPQ]

[The two H's are zinc ligands] [E is the active site residue]

Sequences known to belong to this class detected by the patternALL, except
5 for members of families M5, M7 and M11.

Other sequence(s) detected in SWISS-PROT57; including *Neurospora crassa*
conditiation-specific protein 13 which could be a zinc-protease.

[1]Jongeneel C.V., Bouvier J., Bairoch A. FEBS Lett. 242:211-214(1989).

[2]Murphy G.J.P., Murphy G., Reynolds J.J. FEBS Lett. 289:4-7(1991).

10 [3]Bode W., Grams F., Reinemer P., Gomis-Rueth F.-X., Baumann U., McKay D.B.,
Stoecker W. Zoology 99:237-246(1996).

[4]Rawlings N.D., Barrett A.J. Meth. Enzymol. 248:183-228(1995).

[5]Woessner J. Jr. FASEB J. 5:2145-2154(1991).

[6]Hite L.A., Fox J.W., Bjarnason J.B. Biol. Chem. Hoppe-Seyler 373:381-385(1992).

15 [7]Montecucco C., Schiavo G. Trends Biochem. Sci. 18:324-327(1993).

[8]Niemann H., Blasi J., Jahn R. Trends Cell Biol. 4:179-185(1994).

985. PHO4: Phosphate transporter family

This family includes PHO-4 from *Neurospora crassa* which is a Na(+)-phosphate
20 symporter [1]. This family also contains the leukemia virus receptor Swiss:Q08344. Number
of members: 41

[1] Medline: 95249577 Repressible cation-phosphate symporters in *Neurospora crassa*.
Versaw WK, Metzenberg RL; Proc Natl Acad Sci U S A 1995;92:3884-3887.

986. Photosynthetic reaction center proteins signature (photoRC)

PROSITE cross-reference(s): PS00244; REACTION_CENTER

30 In the photosynthetic reaction center of purple bacteria, two homologous integral
membrane proteins, L(ight) and M(edium), are known to be essential to the light-mediated
water-splitting process. In the photosystem II of eukaryotic chloroplasts two related
proteins are involved: the D1 (psbA) and D2 proteins (psbD). These four types of protein
probably evolved from a common ancestor [see 1,2 for recent reviews].

A signature pattern was developed which include two conserved histidine residues. In L and M chains, the first histidine is a ligand of the magnesium ion of the special pair bacteriochlorophyll, the second is a ligand of a ferrous non-heme iron atom. In photosystem II these two histidines are thought to play a similar role.

Consensus pattern[NQH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2)

[The first H is a magnesium ligand] [The second H is a iron ligand]

Sequences known to belong to this class detected by the patternALL, except for broad bean psbA which has Gln instead of the second His.

[1]Michel H., Deisenhofer J. Biochemistry 27:1-7(1988).

[2]Barber J. Trends Biochem. Sci. 12:321-326(1987).

987. phytochrome: Phytochrome region

This family contains a region specific to phytochrome proteins. Number of members:

145

988. PI3K_C2: C2 domain

Phosphoinositide 3-kinase region postulated to contain a C2 domain. Outlier of C2 family.

Number of members: 39

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.

[2] Medline: 97398940 Phosphoinositide 3-kinases: a conserved family of signal transducers. Vanhaesebroeck B, Leever SJ, Panayotou G, Waterfield MD; Trends Biochem Sci 1997;22:267-272.

989. PI3Ka: Phosphoinositide 3-kinase family, accessory domain (PIK domain)

PIK domain is conserved in all PI3 and PI4-kinases. Its role is unclear but it has been suggested [2] to be involved in substrate presentation.

Number of members: 47

[1] Medline: 97388296 Using structure to define the function of phosphoinositide 3-kinase family members. Domin J, Waterfield MD; FEBS Lett 1997;410:91-95.

[2] Medline: 94069320 Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability. Flanagan CA, Schnieders EA, Emerick AW, Kunisawa R, Admon A,

5 Thorner J; Science 1993;262:1444-1448.

990. P-II protein signatures

PROSITE cross-reference(s): PS00496; PII_GLNB_UMP, PS00638; PII_GLNB_CTER

10 The P-II protein (gene *glnB*) is a bacterial protein important for the control of glutamine synthetase [1,2,3]. In nitrogen-limiting conditions, when the ratio of glutamine to 2-ketoglutarate decreases, P-II is uridylylated on a tyrosine residue to form P-II-UMP. P-II-UMP allows the deadenylation of glutamine synthetase (GS), thus activating the enzyme. Conversely, in nitrogen excess, P-II-UMP is deuridylylated and then promotes the adenylation
15 of GS. P-II also indirectly controls the transcription of the GS gene (*glnA*) by preventing NR-II (*ntrB*) to phosphorylate NR-I (*ntrC*) which is the transcriptional activator of *glnA*. Once P-II is uridylylated, these events are reversed.

20 P-II is a protein of about 110 amino acid residues extremely well conserved. The tyrosine which is uridylylated is located in the central part of the protein.

In cyanobacteria, P-II seems to be phosphorylated on a serine residue rather than being uridylylated.

25 In methanogenic archaeobacteria, the nitrogenase iron protein gene (*nifH*) is followed by two open reading frames highly similar to the eubacterial P-II protein [4]. These proteins could be involved in the regulation of nitrogen fixation.

30 In the red alga, *Porphyra purpurea*, there is a *glnB* homolog encoded in the chloroplast genome.

Other proteins highly similar to *glnB* are:

- *Bacillus subtilis* protein nrgB [5].
- *Escherichia coli* hypothetical protein ybaI [6].

Two signature patterns were developed for P-II protein. The first one is a conserved stretch (in eubacteria) of six residues which contains the uridylylated tyrosine, the other is derived from a conserved region in the C-terminal part of the P-II protein.

Consensus pattern Y-[KR]-G-[AS]-[AE]-Y [The second Y is uridylylated]
Sequences known to belong to this class detected by the pattern ALL glnB's
from eubacteria.

Consensus pattern [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM]

- [1] Magasanik B. *Biochimie* 71:1005-1012(1989).
- [2] Holtel A., Merrick M. *Mol. Gen. Genet.* 215:134-138(1988).
- [3] Cheah E., Carr P.D., Suffolk P.M., Vasudevan S.G., Dixon N.E., Ollis D.L. *Structure* 2:981-990(1994).
- [4] Sibold L., Henriquet M., Possot O., Aubert J.-P. *Res. Microbiol.* 142:5-12(1991).
- [5] Wray L.V. Jr., Atkinson M.R., Fisher S.H. *J. Bacteriol.* 176:108-114(1994).
- [6] Allikmets R., Gerrard B.C., Court D., Dean M.C. *Gene* 136:231-236(1993).

991. PIP5K: Phosphatidylinositol-4-phosphate 5-Kinase

This family contains a region from the common kinase core found in the type I phosphatidylinositol-4-phosphate 5-kinase (PIP5K) family as described in [1]. The family consists of various type I, II and III PIP5K enzymes. PIP5K catalyses the formation of phosphoinositol-4,5-bisphosphate via the phosphorylation of phosphatidylinositol-4-phosphate a precursor in the phosphoinositide signaling pathway. Number of members: 33

- [1] Medline: 98204859. Type I phosphatidylinositol-4-phosphate 5-kinases. Cloning of the third isoform and deletion/substitution analysis of members of this novel lipid kinase family.
- Ishihara H., Shibasaki Y., Kizuki N., Wada T., Yazaki Y., Asano T., Oka Y.; *J Biol Chem* 1998;273:8741-8748.

[2] Medline: 97115834 Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family. Loijens JC, Anderson RA; J Biol Chem 1996 271:32937-32943.

5 992. PolyA_pol: Poly A polymerase family

This family includes nucleic acid independent RNA polymerases, such as Poly(A) polymerase, which adds the poly (A) tail to mRNA EC:2.7.7.19. This family also includes the tRNA nucleotidyltransferase that adds the CCA to the 3' of the tRNA EC:2.7.7.25. Number of members: 31

10

[1] Medline: 93066242 Identification of the gene for an Escherichia coli poly(A) polymerase. Cao GJ, Sarkar N; Proc Natl Acad Sci U S A 1992;89:10380-10384.

993. Photosystem I psaA and psaB proteins signature (psaA_psaB)

15 PROSITE cross-reference(s)PS00419; PHOTOSYSTEM_I_PSAAB

Photosystem I (PSI) [1] is an integral membrane protein complex that uses light energy to mediate electron transfer from plastocyanin to ferredoxin. PSI is found in the chloroplast of plants and cyanobacteria. The electron transfer components of the reaction center of PSI are a primary electron donor P-700 (chlorophyll dimer) and five electron acceptors: A0 (chlorophyll), A1 (a phylloquinone) and three 4Fe-4S iron-sulfur centers: Fx, Fa, and Fb.

20

PsaA and psaB, two closely related proteins, are involved in the binding of P700, A0, A1, and Fx. psaA and psaB are both integral membrane proteins of 730 to 750 amino acids that seem to contain 11 transmembrane segments. The Fx 4Fe-4S iron-sulfur center is bound by four cysteines; two of these cysteines are provided by the psaA protein and the two others by psaB. The two cysteines in both proteins are proximal and located in a loop between the ninth and tenth transmembrane segments. A leucine zipper motif seems to be present [2] downstream of the cysteines and could contribute to dimerization of psaA/psaB.

25

30

The signature pattern for these proteins is based on the perfectly conserved region that includes the two iron-sulfur binding cysteines.

Consensus pattern C-D-G-P-G-R-G-G-T-C [The two C's bind the iron-sulfur center]

[1]Golbeck J.H. Biochim. Biophys. Acta 895:167-204(1987).

[2]Webber A.N., Malkin R. FEBS Lett. 264:1-14(1990).

5 994. PSBH: Photosystem II 10 kDa phosphoprotein

This protein is phosphorylated in a light dependent reaction.

Number of members: 20

995. PsbJ

10 This family consists of the photosystem II reaction center protein PsbJ from plants and Cyanobacteria. In *Synechocystis* sp. PCC 6803 PsbJ regulates the number of photosystem II centers in thylakoid membranes, it is a predicted 4kDa protein with one membrane spanning domain [1]. Number of members: 20

15 [1] Medline: 93131892. Genetic and immunological analyses of the cyanobacterium *Synechocystis* sp. PCC 6803 show that the protein encoded by the *psbJ* gene regulates the number of photosystem II centers in thylakoid membranes. Lind LK, Shukla VK, Nyhus KJ, Pakrasi HB; J Biol Chem 1993;268:1575-1579.

20 996. PSBT: Photosystem II reaction centre T protein

The exact function of this protein is unknown. It probably consists of a single transmembrane spanning helix. The Swiss:P37256 protein, appears to be (i) a novel photosystem II subunit and (ii) required for maintaining optimal photosystem II activity under adverse growth conditions [1]. Number of members: 17

25

[1] Medline: 94298765. The chloroplast *ycf8* open reading frame encodes a photosystem II polypeptide which maintains photosynthetic activity under adverse growth conditions. Monod C, Takahashi Y, Goldschmidt-Clermont M, Rochaix JD; EMBO J 1994;13:2747-2754.

30

997. PSI_8. PHOTOSYSTEM I REACTION CENTRE SUBUNIT VIII. Synonym(s)PSI-I. Gene name(s)PSAI. From *Hordeum vulgare* (Barley). Encoded on Chloroplast. Taxonomy

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

MAY HELP IN THE ORGANIZATION OF THE PSAL SUBUNIT. BELONGS TO THE PSAI FAMILY.

5

[1] SEQUENCE FROM N.A. MEDLINE; 90036933. Scheller H.V., Okkels J.S., Hoej P.B., Svendsen I., Roepstorff P., Moeller B.L.; "The primary structure of a 4.0-kDa photosystem I polypeptide encoded by the chloroplast psaI gene."; J. Biol. Chem. 264:18402-18406(1989).

10 998. PSI_PsaJ: Photosystem I reaction centre subunit IX / PsaJ

This family consists of the photosystem I reaction centre subunit IX or PsaJ from various organisms including *Synechocystis* sp. (strain pcc 6803), *Pinus thunbergii* (green pine) and *Zea mays* (maize). PsaJ Swiss:P19443 is a small 4.4kDa, chloroplastal encoded, hydrophobic subunit of the photosystem I reaction complex its function is not yet fully understood [1].

15

PsaJ can be cross-linked to PsaF Swiss:P12356 and has a single predicted transmembrane domain it has a proposed role in maintaing PsaF in the correct orientation to allow for fast electron transfer from soluble donor proteins to P700+ [1]. Number of members: 18

20

[1] Medline: 99238330. A large fraction of PsaF is nonfunctional in photosystem I complexes lacking the PsaJ subunit. Fischer N, Boudreau E, Hippler M, Drepper F, Haehnel W, Rochaix JD; Biochemistry 1999;38:5546-5552.

[2] Medline: 93252282. Genes encoding eleven subunits of photosystem I from the thermophilic cyanobacterium *Synechococcus* sp. Muhlenhoff U, Haehnel W, Witt H, Herrmann RG; Gene 1993;127:71-78.

25

999. PSII. Protein namePHOTOSYSTEM II P680 CHLOROPHYLL A APOPROTEIN. Synonym(s)CP-47 PROTEIN. Gene name(s)PSBB. From *Hordeum vulgare* (Barley), Encoded on Chloroplast. Taxonomy Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.

30

FUNCTION: THIS PROTEIN CONJUGATES WITH CHLOROPHYLL & CATALYZES THE PRIMARY LIGHT-INDUCED PHOTOCHEMICAL PROCESSES OF PHOTOSYSTEM II. SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE. SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

- [1] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 89240047. Andreeva A.V., Buryakova A.A., Reverdatto S.V., Chakhmakhcheva O.G., Efimov V.A.; "Nucleotide sequence of the 5.2 kbp barley chloroplast DNA fragment, containing psbB-psbH-petB-petD gene cluster."; Nucleic Acids Res. 17:2859-2860(1989).

- [2] SEQUENCE FROM N.A. STRAIN=CV. SABARLIS; MEDLINE; 92207253. Efimov V.A., Andreeva A.V., Reverdatto S.V., Chakhmakhcheva O.G.; "Photosystem II of rye. Nucleotide sequence of the psbB, psbC, psbE, psbF, psbH genes of rye and chloroplast DNA regions adjacent to them."; Bioorg. Khim. 17:1369-1385(1991).

- [3] SEQUENCE OF 411-420. Hinz U.G.; "Isolation of the photosystem II reaction center complex from barley. Characterization by circular dichroism spectroscopy and amino acid sequencing.," Carlsberg Res. Commun. 50:285-298(1985).

1000. QRPTase. Quinolinate phosphoribosyl transferase.

- Quinolate phosphoribosyl transferase (QPRTase) or nicotinate-nucleotide pyrophosphorylase EC:2.4.2.19 is involved in the de novo synthesis of NAD in both prokaryotes and eukaryotes. It catalyses the reaction of quinolinic acid with 5-phosphoribosyl-1-pyrophosphate (PRPP) in the presence of Mg²⁺ to give rise to nicotinic acid mononucleotide (NaMN), pyrophosphate and carbon dioxide [1,2]. Number of members: 26.

- [1] Medline: 97169443. A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. Eads JC, Ozturk D, Wexler TB, Grubmeyer C, Sacchettini JC: *Structure* 1997;5:47-58.

- [2]Medline: 96139309. The sequencing expression, purification, and steady-state kinetic analysis of quinolinate phosphoribosyl transferase from *Escherichia coli*. Bhatia R, Calvo KC; Arch Biochem Biophys 1996;325:270-278.

1001. R3H domain

- The name of the R3H domain comes from the characteristic spacing of the most conserved arginine and histidine residues. The function of the domain is predicted to be binding ssDNA. Number of members: 28

[1]Medline: 99003905 The R3H motif: a domain that binds single-stranded nucleic acids.
Grishin NV; Trends Biochem Sci 1998;23:329-330.

1002. recF protein signatures (RecF)

The prokaryotic protein recF [1,2] is a single-stranded DNA-binding protein which also probably binds ATP. RecF is involved in DNA metabolism; it is required for recombinational DNA repair and for induction of the SOS response. RecF is a protein of about 350 to 370 amino acid residues; there is a conserved ATP-binding site motif 'A' (P-loop) in the N-terminal section of the protein as well as two other conserved regions, one located in the central section, and the other in the C-terminal section. Signature patterns were derived from these two regions.

Consensus pattern [LIVM]-x(4)-[LIF]-x(6)-[LIF]-[LVF]-x-[GE]-[GSTAD]-[PA]- x(2)-R-R-x-[FYW]-[LIVMF]-D Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern[LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L Sequences known to belong to this class detected by the patternALL, except for T. palidum recF.

- [1] Sandler S.J., Chackerian B., Li J.T., Clark A.J. Nucleic Acids Res. 20:839-845(1992).
[2] Alonso J.C., Fisher L.M.; Mol. Gen. Genet. 246:680-686(1995).

1003. RibD C-terminal domain (RibD_C)

The function of this domain is not known, but it is thought to be involved in riboflavin biosynthesis. This domain is found in the C terminus of RibD/RibG Swiss:P25539, in combination with dCMP_cyt_deam, as well as in isolation in some archaeobacterial proteins Swiss:P95872.

Number of members: 21

1004. Ribosomal protein L16 signatures (Ribosomal_L16)

Ribosomal protein L16 is one of the proteins from the large ribosomal subunit. In *Escherichia coli*, L16 is known to bind directly the 23S rRNA and to be located at the A site of the peptidyltransferase center. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial L16.
- Algal and plant chloroplast L16.
- Cyanelle L16.
- Plant mitochondrial L16.

L16 is a protein of 133 to 185 amino-acid residues. As signature patterns, we selected two conserved regions in the central section of these proteins.

Consensus pattern [KR](2)-x-[GSAC]-[KROVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP] Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern R-M-G-x-[GR]-K-G-x(4)-[FWKR] Sequences known to belong to this class detected by the pattern ALL.

[1] Otake E., Hashimoto T., Mizuta K., Suzuki K. *Protein Seq. Data Anal.* 5:301-313(1993).

1005. Ribosomal protein L32e signature (Ribosomal_L32E)

A number of eukaryotic and archaeobacterial ribosomal proteins can be grouped on the basis of sequence similarities. One of these families consists of:

- Mammalian L32 [1].
- *Drosophila* RP49 [2].
- *Trichoderma harzianum* L32 [3].
- Yeast L32e (YBL092w).
- Archaeobacterial L32e [4].

These proteins have 135 to 240 amino-acid residues. As a signature pattern, a stretch of about 20 residues located in the N-terminal part of these proteins was selected.

Consensus pattern F-x-R-x(4)-[KR]-x(2)-[KR]-[LIVMF]-x(3,5)-W-R-[KR]-x(2)-G Sequences known to belong to this class detected by the pattern ALL.

- [1] Jacks C.M., Powaser C.B., Hackett P.B. *Gene* 74:565-570(1988).
 [2] Aguade M. *Mol. Biol. Evol.* 5:433-441(1988).
 [3] Lora J.M., Garcia I., Benitez T., Llobell A., Pintor-Toro J.A. *Nucleic Acids Res.*
 21:3319-3319(1993).
 [4] Arndt E., Scholzen T., Kroemer W., Hatakeyama T., Kimura M. *Biochimie* 73:657-
 668(1991).

1006. (Ribosomal_S3) Ribosomal protein S3 signature

PROSITE: PDOC00474. PROSITE cross-reference(s) PS00548; RIBOSOMAL_S3

Ribosomal protein S3 is one of the proteins from the small ribosomal subunit. In *Escherichia coli*, S3 is known to be involved in the binding of initiator Met-tRNA. It belongs to a family of ribosomal proteins which, on the basis of sequence similarities [1], groups:

- Eubacterial S3.
- Algal and plant chloroplast S3.
- Cyanelle S3.
- Archaeobacterial S3.
- Plant mitochondrial S3.
- Vertebrate S3.
- Insect S3.
- Caenorhabditis elegans* S3 (C23G10.3).
- Yeast S3 (Rp13).

S3 is a protein of 209 to 559 amino-acid residues. A conserved region located in the C-terminal section was selected as a signature pattern.

Consensus pattern[GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-[IDENQ]-x(7)-[LMT]-x(2)-G-x(2)-[GS]. Sequences known to belong to this class detected by the patternALL, except for some mitochondrial S3.

- [1]Otaka E., Hashimoto T., Mizuta K. *Protein Seq. Data Anal.* 5:285-300(1993).

1007. RimM - RimM

The RimM protein is essential for efficient processing of 16S rRNA [1]. The RimM protein was shown to have affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes [1]. Number of members: 14.

- 5 [1]Medline: 98083058. RimM and RbfA are essential for efficient processing of 16S rRNA in *Escherichia coli*. Bylund GO, Wipemo LC, Lundberg LA, Wikstrom PM; J Bacteriol 1998;180:73-82.

1008. RNA_pol_A - RNA polymerase alpha subunit

- 10 -!- RNA polymerases catalyse the DNA dependent polymerisation of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial and chloroplast polymerases).

- 15 -!- Members of this family include: A subunit from eukaryotes, gamma subunit from cyanobacteria, beta' subunit from eubacteria, A' subunit from archaeobacteria, B" from chloroplasts. Number of members: 139.

[1]Medline: 97066998. Structural modules of the large subunits of RNA polymerase. Introducing archaeobacterial and chloroplast split sites in the beta and beta' subunits of *Escherichia coli* RNA polymerase. Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst SA, Goldfarb A; J Biol Chem 1996;271:27969-27974.

1009. RuBisCO_large - Ribulose biphosphate carboxylase large chain active site

PROSITE: PDOC00142; PROSITE cross-reference(s) PS00157; RUBISCO_LARGE

- 25 Ribulose biphosphate carboxylase (EC 4.1.1.39) (RuBisCO) [1,2] catalyzes the initial step in Calvin's reductive pentose phosphate cycle in plants as well as purple and green bacteria. It consists of a large catalytic unit and a small subunit of undetermined function. In plants, the large subunit is coded by the chloroplastic genome while the small subunit is encoded in the nuclear genome. Molecular activation of RuBisCO by CO₂ involves the formation of a carbamate with the epsilon-amino group of a conserved lysine residue. This
- 30 carbamate is stabilized by a magnesium ion. One of the ligands of the magnesium ion is an aspartic acid residue close to the active site lysine [3]. A pattern was developed which includes both the active site residue and the metal ligand, and which is specific to RuBisCO large chains.

Consensus pattern G-x-[DN]-F-x-K-x-D-E [K is the active site residue] [The second D is a magnesium ligand]. Sequences known to belong to this class detected by the pattern ALL, except for *Cheilopleuria bicuspidis* RuBisCO.

[1] Miziorko H.M., Lorimer G.H. *Annu. Rev. Biochem.* 52:507-535(1983).

[2] Akazawa T., Takabe T., Kobayashi H. *Trends Biochem. Sci.* 9:380-383(1984).

[3] Andersson L., Knight S., Schneider G., Lindqvist Y., Lundqvist T., Branden C.-I., Lorimer G.H. *Nature* 337:229-234(1989).

1010. Rve - Integrase core domain

Integrase mediates integration of a DNA copy of the viral genome into the host chromosome. Integrase is composed of three domains. The amino-terminal domain is a zinc binding domain Integrase_Zn. This domain is the central catalytic domain. The carboxyl terminal domain that is a non-specific DNA binding domain integrase. The catalytic domain acts as an endonuclease when two nucleotides are removed from the 3' ends of the blunt-ended viral DNA made by reverse transcription. This domain also catalyses the DNA strand transfer reaction of the 3' ends of the viral DNA to the 5' ends of the integration site [1]. Number of members: 694.

[1] Medline: 95099322. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR; *Science* 1994;266:1981-1986.

1011. (SBP_bac_3) Bacterial extracellular solute-binding proteins, family 3 signature PROSITE: PDOC00798. PROSITE cross-reference(s) PS01039; SBP_BACTERIAL_3

Bacterial high affinity transport systems are involved in active transport of solutes across the cytoplasmic membrane. The protein components of these traffic systems include one or two transmembrane protein components, one or two membrane-associated ATP-binding proteins (ABC transporters; see <PDOC00185>) and a high affinity periplasmic solute-binding protein. The latter are thought to bind the substrate in the vicinity of the inner membrane, and to transfer it to a complex of inner membrane proteins for concentration into the cytoplasm.

In gram-positive bacteria which are surrounded by a single membrane and have therefore no periplasmic region the equivalent proteins are bound to the membrane via an N-terminal lipid anchor. These homolog proteins do not play an integral role in the transport process per se, but probably serve as receptors to trigger or initiate translocation of the solute through the membrane by binding to external sites of the integral membrane proteins of the efflux system.

In addition at least some solute-binding proteins function in the initiation of sensory transduction pathways.

On the basis of sequence similarities, the vast majority of these solute-binding proteins can be grouped [1] into eight families of clusters, which generally correlate with the nature of the solute bound.

Family 3 groups together specific amino acids and opine-binding periplasmic proteins and a periplasmic homolog with catalytic activity:

-Histidine-binding protein (gene *hisJ*) of *Escherichia coli* and related bacteria. An homologous lipoprotein exists in *Neisseria gonorrhoeae*.

-Lysine/arginine/ornithine-binding proteins (LAO) (gene *argT*) of *Escherichia coli* and related bacteria are involved in the same transport system than *hisJ*. Both solute-binding proteins interact with a common membrane-bound receptor *hisP* of the binding protein dependent transport system *HisQMP*.

-Glutamine-binding proteins (gene *glnH*) of *Escherichia coli* and *Bacillus stearothermophilus*.

-Glutamate-binding protein (gene *gluB*) of *Corynebacterium glutamicum*.

-Arginine-binding proteins *artI* and *artJ* of *Escherichia coli*.

-Nopaline-binding protein (gene *nocT*) from *Agrobacterium tumefaciens*.

-Octopine-binding protein (gene *occT*) from *Agrobacterium tumefaciens*.

-Major cell-binding factor (CBF1) (gene: *peb1A*) from *Campylobacter jejuni*.

-*Bacteroides nodosus* protein *aabA*.

-Cyclohexadienyl/arogenate dehydratase of *Pseudomonas aeruginosa*, a periplasmic enzyme which forms an alternative pathway for phenylalanine biosynthesis.

-*Escherichia coli* protein *fliY*.

-*Vibrio harveyi* protein *patH*.

-*Escherichia coli* hypothetical protein *ydhW*.

-*Bacillus subtilis* hypothetical protein *yckB*.

-Bacillus subtilis hypothetical protein yckK.

The signature pattern is located near the N-terminus of the mature proteins.

Consensus patternG-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-
5 [LIVMAGN]

Sequences known to belong to this class detected by the patternALL.

[1]Tam R., Saier M.H. Jr. Microbiol. Rev. 57:320-346(1993).

10 1012. Sec7 - Sec7 domain

The Sec7 domain is a guanine-nucleotide-exchange-factor (GEF)for the arf family [2].

Number of members: 32.

[1]Medline: 98169075. Structure of the Sec7 domain of the Arf exchange factor. ARNO.

15 Cherfils J, Menetrey J, Mathieu M, Le Bras G, Robineau S, Beraud-Dufour S, Antonny B, Chardin P; Nature 1998;392:101-105.

[2]Medline: 97100951. A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. Chardin P, Paris S, Antonny B, Robineau S, Beraud-Dufour S, Jackson CL, Chabre M. Nature 1996;384:481-484.

20 1013. SecA_{protein}. SecA protein, amino terminal region

SecA protein binds to the plasma membrane where it interacts with proOmpA to support translocation of proOmpA through the membrane. SecA protein achieves this translocation, in association with SecY protein, in an ATP dependent manner. SecA possesses the ATPase
25 activity. The carboxyl terminus has similarity with the helicase carboxyl terminus. See Ribosomal_L5. Number of members: 45.

[1]Medline: 98309858. Amino-terminal region of SecA is involved in the function of SecE for protein translocation into Escherichia coli membrane vesicles. Mori H, Sugiyama H,

30 Yamanaka M, Sato K, Tagaya M, Mizushima S; J Biochem (Tokyo) 1998;124:122-129.

[2]Medline: 89251629. SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. Lill R, Cunningham K, Brundage LA, Ito K, Oliver D, Wickner W; EMBO J 1989;8:961-966.

1014. Seedstore_2S - 2S seed storage family

Members of this family are composed of two chains (both included in the alignment), these are co-translated and later cleaved. The two chains are disulphide linked together. Number of members: 27.

[1]Medline: 97121264. 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. Rico M, Bruix M, Gonzalez C, Monsalve RI, Rodriguez R; Biochemistry 1996;35:15672-15682.

1015. Smr - Smr domain

This family includes the Smr (Small MutS Related) proteins, and the C-terminal region of the MutS2 protein. It has been suggested that this domain interacts with the MutS1 Swiss:P23909 protein in the case of Smr proteins and with the N-terminal MutS related region of MutS2 Swiss:P94545 [1]. Number of members: 14.

[1]Medline: 10431172. Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. Moreira D, Philippe H; Trends Biochem Sci 1999;24:298-300.

1016. (SSF) Sodium:solute symporter family signatures and profile

PROSITE: PDOC00429. PROSITE cross-reference(s)PS00456; NA_SOLUT_SYMP_1 PS00457; NA_SOLUT_SYMP_2 PS50283; NA_SOLUTE_SYMP_3

It has been shown [1,2] that integral membrane proteins that mediate the intake of a wide variety of molecules with the concomitant uptake of sodium ions (sodium symporters) can be grouped, on the basis of sequence and functional similarities into a number of distinct families. One of these families is known as the sodium:solute symporter family (SSF) and currently consists of the following proteins:

- Mammalian Na⁺/glucose co-transporter.
- Mammalian Na⁺/myo-inositol co-transporter.
- Mammalian Na⁺/nucleoside co-transporter.
- Mammalian Na⁺/neutral amino acid co-transporter.
- Escherichia coli Na⁺/proline symporter (gene putP).
- Escherichia coli Na⁺/pantothenate symporter (gene panF).

- Escherichia coli hypothetical protein yidK.
- Escherichia coli hypothetical protein yjcG.
- Bacillus subtilis hypothetical protein ywca (ipa-31R).

These integral membrane proteins are predicted to comprise at least ten membrane spanning domains. Two conserved regions were selected as signature patterns; the first one is located in the fourth transmembrane region and the second one in a loop between two transmembrane regions in the C-terminal part of these proteins.

Consensus pattern[GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-[SAP]. Sequences known to belong to this class detected by the patternALL.

Consensus pattern[GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-x-[LIVMGA] Sequences known to belong to this class detected by the patternALL, except for E.coli yidK.

Note this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.

[1]Reizer J., Reizer A., Saier M.H. Jr. Res. Microbiol. 141:1069-1072(1991).

[2]Reizer J., Reizer A., Saier M.H. Jr. Biochim. Biophys. Acta 1197:133-136(1994).

1017. SurE - Survival protein SurE

E. coli cells with the surE gene disrupted are found to survive poorly in stationary phase [1]. It is suggested that SurE may be involved in stress response. Yeast also contains a member of the family Swiss:P38254. Swiss:P30887 can complement a mutation in acid phosphatase, suggesting that members of this family could be phosphatases. Number of members: 17.

[1]Medline: 95014035. A new gene involved in stationary-phase survival located at 59 minutes on the Escherichia coli chromosome. Li C, Ichikawa JK, Ravetto JJ, Kuo HC, Fu JC, Clarke S; J Bacteriol 1994;176:6015-6022.

[2]Medline: 93046805. Complementation of Saccharomyces cerevisiae acid phosphatase mutation by a genomic sequence from the yeast Yarrowia lipolytica identifies a new phosphatase. Treton BY, Le Dall MT, Gaillardin CM; Curr Genet 1992;22:345-355.

1018. Synuclein - Synuclein

There are three types of synucleins in humans, these are called alpha, beta and gamma. Alpha synuclein has been found mutated in families with autosomal dominant Parkinson's disease. A peptide of alpha synuclein has also been found in amyloid plaques in Alzheimer's patients. Number of members: 12.

[1]Medline: 98424410. The synuclein family. Lavedan C; Genome Res 1998;8:871-880.

1019. (T-box) T-box domain signatures

PROSITE: PDOC00972. PROSITE cross-reference(s) PS01283; TBOX_1 PS01264; TBOX_2

A number of eukaryotic DNA-binding proteins contain a domain of about 170 to 190 amino acids known as the T-box domain [1,2,3] and which probably binds DNA. The T-box has first been found in the mice T locus (Brachyury) protein, a transcription factor involved in mesoderm differentiation. It has since been found in the following proteins:

- Vertebrate and invertebrate homologs of the T protein.
- Mammalian proteins TBX1 to TBX6.
- Mammalian protein TBR1 which is expressed specifically in brain.
- Xenopus laevis eomesodermin (eomes).
- Xenopus laevis VegT (or Antipodean), a transcription factor that activates the expression of wnt-8, eomes and Brachyury.
- Chicken TbxT.
- Drosophila protein optomotor-blind (omb).
- Drosophila protein brachyenteron (byn) (also known as Trg), which is required for the specification of the hindgut and anal pads.
- Drosophila protein H15.
- Caenorhabditis elegans protein tbx-12.
- Caenorhabditis elegans hypothetical proteins F21H11.3, F40H6.4, T07C4.2, T07C4.6 and ZK177.10.

Two conserved regions were selected as signature patterns for the T-domain. The first region corresponds to the N-terminal of the domain and the second one to the central part.

Consensus pattern L-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ]

Sequences known to belong to this class detected by the patternALL, except for *C.elegans* ZK177.10.

Consensus pattern[LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x- F

Sequences known to belong to this class detected by the patternALL, except for *C.elegans* tbx-12, ZK177.10 and *Drosophila* H15.

[1]Bollag R.J., Siegfried Z., Cebra-Thomas J.A., Garvey N., Davison E.M., Silver L.M. Nat. Genet. 7:383-389(1994).

[2] Agulnik S.I., Garvey N., Hancock S., Ruvinsky I., Chapman D.L., Agulnik I., Bollag R.J., Papaioannou V.E., Silver L.M. Genetics 144:249-254(1996).

[3]Papaioannou V.E. Trends Genet. 13:212-213(1997).

1020. Toprim - Toprim domain

This is a conserved region from DNA primase. This corresponds to the Toprim domain common to DnaG primases, topoisomerases, OLD family nucleases and RecR proteins [1]. Both DnaG motifs IV and V are present in the alignment, the DxD (V) motif may be involved in Mg2+ binding and mutations to the conserved glutamate (IV) completely abolish DnaG type primase activity [1]. DNA primase EC:2.7.7.6 is a nucleotidyltransferase it synthesizes the oligoribonucleotide primers required for DNA replication on the lagging strand of the replication fork; it can also prime the leading stand and has been implicated in cell division [2]. Number of members: 133.

[1]Medline: 98391745. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Aravind L, Leipe DD, Koonin EV; Nucleic Acids Res 1998;26:4205-4213.

[2]Medline: 97368180. Cloning and analysis of the dnaG gene encoding *Pseudomonas putida* DNA primase. Szafranski P, Smith CL, Cantor CR; Biochim Biophys Acta 1997;1352:243-248.

[3]Medline: 94124015. The *Haemophilus influenzae* dnaG sequence and conserved bacterial primase motifs. Versalovic J, Lupski JR; Gene 1993;136:281-286.

1021. TraB - TraB family

pAD1 is a hemolysin/bacteriocin plasmid originally identified in *Enterococcus faecalis* DS16.

It encodes a mating response to a peptide sex pheromone, cAD1, secreted by recipient bacteria. Once the plasmid pAD1 is acquired, production of the pheromone ceases--a trait related in part to a determinant designated traB. However a related protein is found in *C.*

- 5 elegans Swiss:Q94217, suggesting that members of the TraB family have some more general function. Number of members: 12.

[1]Medline: 94302142. Characterization of the determinant (traB) encoding sex pheromone shutdown by the hemolysin/bacteriocin plasmid pAD1 in *Enterococcus faecalis*. An FY,
10 Clewell DB; Plasmid 1994;31:215-221.

1022. (Transpo_mutator) Transposases, Mutator family, signature
PROSITE: PDOC00770. PROSITE cross-reference(s) PS01007;
TRANSPOSASE_MUTATOR

15 Autonomous mobile genetic elements such as transposon or insertion sequences (IS) encode an enzyme, called transposase, required for excising and inserting the mobile element. On the basis of sequence similarities, transposases can be grouped into various families. One of these families has been shown [1,2,3,E1] to consist of transposases from the following elements:

- 20 -Mutator from Maize.
-Is1201 from *Lactobacillus helveticus*.
-Is905 from *Lactococcus lactis*.
-Is1081 from *Mycobacterium bovis*.
-Is6120 from *Mycobacterium smegmatis*.
25 -Is406 from *Pseudomonas cepacia*.
-IsRm3 from *Rhizobium meliloti*.
-IsRm5 from *Rhizobium meliloti*.
-Is256 from *Staphylococcus aureus*.
-IsT2 from *Thiobacillus ferrooxidans*.

- 30 The maize Mutator transposase (MudrA) is a protein of 823 amino acids; the bacterial transposases listed above are proteins of 300 to 420 amino acids. These proteins contain a conserved domain of about 130 residues; a signature pattern was derived from the most conserved part of this domain.

- 5 [1]Eisen J.A., B  nito M.-I., Walbot V. Nucleic Acids Res. 22:2634-2636(1994).
[2]Guillhot C., Gicquel B., Davies J., Martin C. Mol. Microbiol. 6:107-113(1992).
[3]Wood M.S., Byrne A., Lessie T.G. Gene 105:101-105(1991).

1023. Transposase 8 - Transposase

10 Transposase proteins are necessary for efficient DNA transposition. This family consists of various E. coli insertion elements and other bacterial transposases some of which are members of the IS3 family. Number of members: 58.

- [1]Medline: 97324595. Genetic organization and transposition properties of IS511. D. A. Mullin, D. L. Zies, A. H. Mullin, N. Caballera & B. Ely; Mol Gen Genet 1997;254:456-463.
- [2]Medline: 97128810. The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of *Streptomyces lividans* 66. J. Fischer, H. Maier, P. Viell & J. Altenbuchner; Gene 1996;180:81-89.
- [3]Medline: 97074647. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence IS_{Rm6}, a small transposable element that belongs to the IS3 family. S. Zekri & N. Toro; Gene 1996;175:43-48.

1024. tRNA int endo - tRNA intron endonuclease

Members of this family cleave pre tRNA at the 5' and 3' splice sites to release the intron

25 EC:3.1.27.9. Number of members: 8.

- [1]Medline: 97344075. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Kleman-Lever K, Armbruster DW, Daniels CJ; Cell 1997;89:839-847.

1025. Urease - Urease signatures

PROSITE: PDOC00133PROSITE cross-reference(s) PS01120; UREASE_1 PS00145;
UREASE 2

Urease (EC 3.5.1.5) is a nickel-binding enzyme that catalyzes the hydrolysis of urea to carbon dioxide and ammonia [1]. Historically, it was the first enzyme to be crystallized (in 1926). It is mainly found in plant seeds, microorganisms and invertebrates. In plants, urease is a hexamer of identical chains. In bacteria [2], it consists of either two or three different subunits (alpha, beta and gamma).

Urease binds two nickel ions per subunit; four histidine, an aspartate and a carbamated-lysine serve as ligands to these metals; an additional histidine is involved in the catalytic mechanism [3].

As signatures for this enzyme, a region that contains two histidine that bind one of the nickel ions and the region of the active site histidine was selected.

Consensus pattern T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P [The two H's bind nickel]. Sequences known to belong to this class detected by the pattern ALL.

Consensus pattern [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A [H is the active site residue]. Sequences known to belong to this class detected by the pattern ALL.

[1] Takishima K., Suga T., Mamiya G. Eur. J. Biochem. 175:151-165(1988).

[2] Mobley H.L.T., Husinger R.P. Microbiol. Rev. 53:85-108(1989).

[3] Jabri E., Carr M.B., Hausinger R.P., Karplus P.A. Science 268:998-1004(1995).

1026. Urease_beta - Urease beta subunit.

This subunit is known as alpha in *Helicobacter*. Number of members: 35.

[1] Medline: 95273988. The crystal structure of urease from *Klebsiella aerogenes*. Jabri E,

Carr MB, Hausinger RP, Karplus PA; Science 1995;268:998-1004.

1027. UvrD-helicase - UvrD/REP helicase

The Rep family helicases are composed of four structural domains. The Rep family function as dimers. REP helicases catalyze ATP dependent unwinding of double stranded DNA to single stranded DNA. Swiss:P23478, Swiss:P08394 have large insertions near to the carboxy-terminus relative to other members of the family. Number of members: 52.

[1] Medline: 97433075. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G; Cell 1997;90:635-647.

5 1028. V-type ATPase 116kDa subunit family (V_ATPase_sub_a)

This family consists of the 116kDa V-type ATPase (vacuolar (H⁺)-ATPases) subunits, as well as V-type ATP synthase subunit i. The V-type ATPases family are proton pumps that acidify intracellular compartments in eukaryotic cells for example yeast central vacuoles, clathrin-coated and synaptic vesicles. They have important roles in membrane trafficking processes [1]. The 116kDa subunit (subunit a) in the V-type ATPase is part of the V0 functional domain responsible for proton transport. The a subunit is a transmembrane glycoprotein with multiple putative transmembrane helices. It has a hydrophilic amino terminal and a hydrophobic carboxy terminal [1,2]. It has roles in proton transport and assembly of the V-type ATPase complex [1,2]. This subunit is encoded by two homologous gene in yeast VPH1 and STV1 [2].

Number of members: 27

[1] Forgac M; Medline: 99240666 "Structure and properties of the vacuolar (H⁺)-ATPases." J Biol Chem 1999;274:12951-12954.

[2] Forgac M; Medline: 99270697 "Structure and properties of the clathrin-coated vesicle and yeast vacuolar V-ATPases." J Bioenerg Biomembr 1999;31:57-65.

1029. Viral (Superfamily 1) RNA helicase (Viral_helicase1)

Number of members: 260

[1] Koonin EV, Dolja VV; Medline: 94094568 "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." Crit Rev Biochem Mol Biol 1993;28:375-430.

1030. Vesicular monoamine transporter (VMAT)

This family consists of various vesicular amine transporters with 12 transmembrane helices. These included vesicular acetylcholine transporters (VACHT) [3], and vesicular monoamine transporters (VMATs) [1,2] isoforms 1 adrenal and 2 brain (VMAT1 and VMAT2).

These proteins transport biogenic amines into synaptic vesicles or chromaffin granules [4]. VMATs pack monoamine neurotransmitters into secretory vesicles for regulated exocytotic release, they also protect against the parkinsonian neurotoxins MPP+ by transporting it into vesicles preventing it from acting on mitochondria [1].

Also in the family is *C. elegans* UNC-17 a putative vesicular acetylcholine transporter mutations in UNC-17 cause impaired neuromuscular function, giving rise to jerky or uncoordinated movement, [4].

Number of members: 15

[1] Krantz DE, Peter D, Liu Y, Edwards RH; Medline: 97197857 "Phosphorylation of a vesicular monoamine transporter by casein kinase II." *J Biol Chem* 1997;272:6752-6759.

[2] Erickson JD, Varoqui H, Schafer MK, Modi W, Diebler MF, Weihe E, Rand J, Eiden LE, Bonner TI, Usdin TB; Medline: 94350930 "Functional identification of a vesicular acetylcholine transporter and its expression from a 'cholinergic' gene locus." *J Biol Chem* 1994;269:21929-21932.

[3] Erickson JD, Schafer MK, Bonner TI, Eiden LE, Weihe E; Medline: 96209876 "Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter." *Proc Natl Acad Sci U S A* 1996;93:5166-5171.

[4] Alfonso A, Grundahl K, Duerr JS, Han HP, Rand JB; Medline: 3342494 "The *Caenorhabditis elegans* unc-17 gene: a putative vesicular acetylcholine transporter." *Science* 1993;261:617-619.

1031. WW/rsp5/WWP domain signature and profile. Cross-reference(s): PS01159; WW_DOMAIN_1; PS50020; WW_DOMAIN_2

The WW domain [1-4,E1] (also known as rsp5 or WWP) has been originally discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues,

is repeated up to 4 times in some proteins. It has been shown [5] to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain are listed below.

- 10 --Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophin form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.
- 15 --Utrophin, a dystrophin-like protein of unknown function.
- Vertebrate YAP protein is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains [6].
- 20 --Mouse NEDD-4 plays a role in the embryonic development and differentiation of the central nervous system. It contains 3 WW modules followed by a HECT domain. The human ortholog contains 4 WW domains, but the third WW domain is probably spliced resulting in an alternate NEDD-4 protein with only 3 WW modules [3].
- 25 --Yeast RSP5 is similar to NEDD-4 in its molecular organization. It contains an N-terminal C2 domain (see <PDOC00380>), followed by a histidine-rich region, 3 WW domains and a HECT domain.
- 30 --Rat FE65, a transcription-factor activator expressed preferentially in liver. The activator domain is located within the N-terminal 232 residues of FE65, which also contain the WW domain.

--Yeast ESS1/PTF1, a putative peptidyl prolyl cis-trans isomerase from family ppiC (see <PDOC00840>). A related protein, dodo (gene dod) exists in *Drosophila* and in mammals (gene PIN1).

--Tobacco DB10 protein. The WW domain is located N-terminal to the region with similarity to ATP-dependent RNA helicases.

--IQGAP, a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

--Yeast pre-mRNA processing protein PRP40, *Caenorhabditis elegans* ZK1098.1 and fission yeast SpAC13C5.02 are related proteins with similarity to MYO2-type myosin, each containing two WW-domains at the N-terminus.

--*Caenorhabditis elegans* hypothetical protein C38D4.5, which contains one WW module, a PH domain (see <PDOC50003>) and a C-terminal phosphatidylinositol 3-kinase domain.

--Yeast hypothetical protein YFL010c.

For the sensitive detection of WW domains, a profile was developed which spans the whole homology region as well as a pattern.

Description of pattern(s) and/or profile(s):

Consensus pattern W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

[1] Bork P., Sudol M. Trends Biochem. Sci. 19:531-533(1994).

[2] Andre B., Springael J.Y. Biochem. Biophys. Res. Commun. 205:1201-1205(1994).

[3] Hofmann K.O., Bucher P. FEBS Lett. 358:153-157(1995).

[4] Sudol M., Chen H.I., Bougeret C., Einbond A., Bork P. FEBS Lett. 369:67-71(1995).

[5] Chen H.I., Sudol M. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823(1995).

[6] Sudol M., Bork P., Einbond A., Kastury K., Druck T., Negrini M., Huebner K., Lehman D. J. Biol. Chem. 270:14733-14741(1995).

1032. XPA protein signatures. cross-reference(s): XPA_1 PROSITE PS00752; PS00753; XPA_2.

Xeroderma pigmentosum (XP) [1] is a human autosomal recessive disease, characterized by a high incidence of sunlight-induced skin cancer. People's

skin cells with this condition are hypersensitive to ultraviolet light, due to defects in the incision step of DNA excision repair. There are a minimum of seven genetic complementation groups involved in this pathway: XP-A to XP-G. XP-A is the most severe form of the disease and is due to defects in a 30 Kd nuclear protein called XPA (or XPAC) [2].

The sequence of the XPA protein is conserved from higher eukaryotes [3] to yeast (gene RAD14) [4]. XPA is a hydrophilic protein of 247 to 296 amino-acid residues which has a C4-type zinc finger motif in its central section.

Two signature were developed patterns for XPA proteins. The first corresponds to the zinc finger region, the second to a highly conserved region located some 12 residues after the zinc finger region.

Consensus pattern C-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C
Consensus pattern [LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE]

[1] Tanaka K., Wood R.D. Trends Biochem. Sci. 19:83-86(1994).

[2] Miura N., Miyamoto I., Asahina H., Satokata I., Tanaka K., Okada Y. J. Biol. Chem. 266:19786-19789(1991).

[3] Shimamoto T., Kohno K., Tanaka K., Okada Y. Biochem. Biophys. Res. Commun. 181:1231-1237(1991).

[4] Bankmann M., Prakash L., Prakash S. Nature 355:555-558(1992).

1033. YCF9

This family consists of the hypothetical protein product of the YCF9 gene from chloroplasts and cyanobacteria. Number of members: 16

1034. (DUF15)

It is highly conserved between eubacteria and eukaryotes.

Number of members: 30

1035. Luminal portion of Cytochrome b559, alpha (gene psbE) subunit. (cytochr_b559a)

This family is the luminal portion of cytochrome b559 alpha chain, matches to this family
should be accompanied by a match to the cytochr_b559 family also. The Prosite pattern
pattern matches the transmembrane region of the cytochrome b559 alpha and beta subunits.
Number of members: 16

A. Asparaginase 2

Asparaginase II (L-asparagine aminohydrolase II) is an extracellular protein that may be associated with the cell wall and whose expression is affected by the availability of nitrogen.

Asparaginase II catalyzes the reaction of L-Asparagine + H₂O = L-Aspartate + NH₃. As many leukemias have high requirements for aspartic acid, asparaginase II proteins are useful as reagents for screening compounds for activity as leukemia chemotherapy products. Asparaginase II protein can also be over- or under-expressed to alter amino acid content in plant tissues or to modify nitrogen fixation and/or nitrogen metabolism in plants.

Ref: Bon et al. (1997) Appl Biochem Biotechnol 63-65: 203-12

B. Chloroa b-bind

Chlorophyll a-b binding proteins are located in the thylakoid membranes of the chloroplast and bind chlorophyll a and chlorophyll b, thereby triggering a chemical reaction (photosynthesis). These proteins are useful in controlling the rate, efficiency and/or output of photosynthesis. Overexpression of chlorophyll a-b binding proteins is expected to increase the rate of photosynthesis.

Ref: Leutwiler et al. (1986) Nucleic Acids Res 14: 4051-64

Brandt et al. (1992) Plant Mol Biol 19: 699-703

C. DMRL synthase

DMRL Synthase (6,7-Dimethyl-8-Ribityllumazine Synthase) catalyzes the last step in riboflavin (Vitamin B₂) synthesis, condensing 5-amino-6-(1'-D)-ribityl-amino-2,4(1H, 3H)-Pyrimidinedione with L-3,4-Dihydroxy-2-Butanone 4-Phosphate producing 6,7-Dimethyl-8-(1-D-Ribityl)Luminazine . The enzyme forms a homopentamer. Engineering of these proteins or those with homologous sequences/structures may allow control of the amounts of vitamin B₂ available in plants and/or accumulation of pigment, as well as altering reactions requiring hydrogen ion carriers/transmitters.

Ref: Garcia-Ramirez et al. (1995) J Biol Chem **270**: 23801-7

D. E1_N

These proteins are ATP-dependent DNA helicases that are required for initiation of viral DNA replication. They form a complex with the viral E2 protein. The E1-E2 complex binds to the replication origin that contains binding sites for both proteins. The majority of sequences known for this group of proteins are from various papillomaviruses, a type of double stranded DNA virus. In plants, the prototype double stranded DNA virus is Cauliflower Mosaic virus (CaMV). Manipulation of these proteins, especially to produce variant proteins that form non-productive complexes, enables production of plants that are resistant to infection by double stranded DNA viruses.

Ref: Yang et al. (1993) PNAS USA **90**: 5086-90

Ustav and Stenlund (1991) EMBO J **10**: 449-57

Callaway et al. (1996) Mol Plant Microbe Interact **9**: 810-8

E. EF1_G

Elongation Factor-1 is composed of four subunits: alpha, beta, delta and gamma. Gamma subunits are presumed to play a role in anchoring the complex to other cellular components. Studies of EF-1 genes in plants suggests that different forms of the EF-1 subunits may be expressed in particular organs or in response to stress. Manipulation of the activity of these

proteins, either by altered expression level or by structural mutation, may result in the accumulation of a particular protein in a chosen organ or allow production of particular proteins during stress conditions.

- 5 Ref: Kinzy et al. (1994) NAR 22: 2703-7
 Dunn et al. (1993) Plant Mol Biol 23: 221-5
 Aguilar et al. (1991) Plant Mol Biol 17: 351-60

F. ENV polyprotein

10 This family comprises the envelope or coat proteins known from a number of different retroviruses. In mammalian species, retroviruses are responsible for diseases such as leukemia and HIV. In plants, retroviruses are known in both monocot (e.g. Zeon-1) and dicot (e.g. Arabidopsis and tobacco) species and have been shown to induce mutant alleles at new
 15 loci. Engineering of plant ENV proteins may allow mobilization or targeting of endogenous or introduced retroviruses, in essence generating a new method for mutant production, gene tagging and the like.

- 20 Ref: Mamoun et al (1990) J Virol 64: 4180-8
 Grandbastien et al. (1989) Nature 337: 376-80
 Wright and Voytas (1998) Genetics 149: 703-15

G. Glycosyl_hydr9

25 Proteins having this domain (previously known as the glycosyl hydrolase family 5 domain) catalyze the endohydrolysis of 1,4- β -D-glucosidic linkages in cellulose. Numerous plant proteins with this domain exist and are expressed in an organ specific manner. They are involved in the fruit ripening process, in cell elongation and plant reproduction. Modulation
 30 of the activity of these proteins, either by over- or under-expression or by mutation of the polypeptide, could be used to affect post-harvest physiology (e.g. rate of ripening) or for engineering reproductive sterility.

- Ref: Giorda et al. (1990) *Biochemistry* 29: 7264-9
Tucker et al. (1988) *Plant Physiol* 88: 1257-62
Shani et al. (1997) 43: 837-42
Milligan and Gasser (1995) *Plant Mol Biol* 28: 691-711

5

H. Glycosyl_hydr14

The β -amylases (family 14 of glycosyl hydrolases) catalyze the hydrolysis of 1,4- α -glucosidic linkages in polysaccharides and remove successive maltose units from the non-reducing ends of the chains. Mutants of β -amylase in *Arabidopsis* exhibited altered degradation of starch throughout the diurnal cycle. In addition, the mutant phenotypes indicated that these enzymes not only affect carbohydrate metabolism/catabolism, but also influence the amount of pigment stored within particular cells. Manipulation of the β -amylase genes enables control of plant pigmentation (for example, fibre pigment in cotton) as well as carbohydrate synthesis and degradation.

10

15

- Ref: Zeeman et al. (1998) *Plant J* 15: 357-65
Hirano and Nakamura (1997) *Plant Physiol* 114: 5675-82
Kitamoto et al. (1988) *J Bacteriol* 170: 5848-54

20

I. Glycosyl_hydr15

Glycosyl hydrolases from family 15 (such as 1,4-Alpha-D-Glucan glucohydrolase,) catalyze the hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from the non-reducing ends of the chains resulting in the release of β -D-Glucose. In plants these proteins have been tied to the mobilization of the xyloglucan stored in the cotyledonary cell walls. Proteins such as these could be varied to affect the rate of plant growth (for example during germination), storage and/or use of glucose and other sugars by plant tissues and alteration of the properties, such as elasticity, of plant cell walls.

25

30

- Ref: Crombie et al. (1998) *Plant J* 15: 27-38
Hata et al. (1991) *Agric Biol Chem* 55: 941-9

J. Glycosyl_hydr20

Members of the family 20 glycosyl hydrolases catalyze the hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl- β -D-hexosaminides. N-acetyl- β -glucosaminidase belongs to this family and exists in several different forms (consisting of various combinations of alpha and beta chains) depending on the organism. Family 20 glycosyl hydrolases have been implicated in lysosomal storage diseases (such as Sandhoff disease) and glycogen storage disease in humans. These types of proteins are also responsible for the hydrolysis of chitin. In plants, these proteins could be useful in controlling carbohydrate catabolism, thereby influencing the amount of sugars available for storage and/or use in other metabolic pathways. In addition, it is possible that such proteins could be used to engineer an endogenous insect protection mechanism, e.g. by secretion of a chitin-hydrolyzing composition by the plant.

Ref: Graham et al (1988) J Biol Chem 263: 16823-9
O'Dowd et al. (1988) Biochemistry 27: 5216-26

K. HMG box

The HMG box is a novel type of DNA-binding domain found in a diverse group of proteins. Numerous plant proteins contain this domain, such as the HMG1/2-like proteins. The expression of some of these HMG proteins appears to be regulated by circadian rhythms and in a light dependent manner, occurring at higher levels in roots, for example and lower levels in light-grown tissues such as cotyledons. Generally, HMG proteins are thought to influence transcription regulation. In plants, HMGs are believed to have a role in maintaining patterns of circadian-regulated expression for other genes, suggesting that these proteins could be exploited to control growth and development.

Ref: Laudet et al. (1993) Nucleic Acids Res 21: 2493-501
Zheng et al. (1993) Plant Mol Biol 23: 813-23
Grasser et al. (1993) Plant Mol Biol 23: 619-25

L. IL2

Interleukin-2 (IL-2) is produced in mammals by T cells in response to antigenic or mitogenic stimulation and is crucial for proper regulation and functioning of the immune response. IL-2 is capable of stimulating B cells, monocytes, lymphokine-activated killer cells, natural killer cells and glioma cells. Plant extracts have also been shown to stimulate the immune system (for example, mistletoe therapy for human cancer). It is known that IL-2 is involved in feedback inhibition pathways that impact the inflammatory response as well as the growth inhibition of tumor reactive T cells. Plant proteins containing IL-2-like sequences are useful as immunity-based therapeutics, acting in a manner similar to IL-2 in mammals.

- Ref: Heike et al. (1997) Scand J Immunol 45: 221-6
Ariel et al. (1998) J Immunol 161: 2465-72
Schink (1997) Anticancer Drugs 8 Suppl 1: S47-51

M. Oxidored_FMN

NADPH dehydrogenases catalyze the reaction $\text{NADPH} + \text{acceptor} = \text{NADP}(+) + \text{reduced acceptor}$. One member of this family is yeast "old yellow enzyme" (OYE) and is thought to be involved in oxylipin metabolism. A second yeast family member is a protein that binds estrogen binding protein (EBP) in addition to exhibiting oxidoreductase activity. An Arabidopsis homolog to OYE has been described and estrogen binding proteins in plants have been reported. Plant proteins from this class have the potential to be used to modify lipid metabolism/catabolism. These proteins may also have use as therapeutics for breast and prostate cancer, and other abnormal growth in steroid-sensitive tissues.

- Ref: Baker et al. (1998) Proc Soc Exp Biol Med 217: 317-21
Schaller and Weiler (1997) J Biol Chem 272: 28066-72
Mandani et al. (1994) PNAS USA 91: 922-6

N. Oxidored_q2

The NADH-plastoquinone oxidoreductases catalyze the reaction $\text{NADH} + \text{plastoquinone} = \text{NAD}(+) + \text{plastoquinol}$. In plants these reactions occur in the chloroplast and are believed to participate in a chloroplast respiratory system. Here, the NDH complex is postulated to act as a valve to remove excess reduction equivalents in the chloroplasts. Manipulation of these proteins may improve the rate or efficiency of photosynthesis.

Ref: Burrows et al. (1998) EMBO J 17: 868-76

Kofer et al (1998) Mol Gen Genet 258: 166-73

Maier et al. (1995) J Mol Biol 251: 614-28

O. PABP

Polyadenylate binding proteins bind the poly (A) tail of mRNA. Plants, as exemplified by Arabidopsis, contain numerous PABP genes that are expressed in an organ-specific manner. For example, PABP2 is functional in roots and shoots, while PABP5 is expressed predominantly in immature flowers. The PABP proteins are implicated in numerous aspects of posttranscriptional regulation including mRNA turnover and translational initiation. Control of activity of PABP proteins provides the ability to control the expression of various genes in particular organs during development.

Ref: Hilson et al (1993) Plant Physiol 103: 525-33

Belostotsky and Meagher (1993) PNAS USA 90: 6686-90

P. Parvo coat

Parvoviruses are linear single-stranded DNA viruses that are encapsulated by three capsid proteins. Plants are susceptible to infection by single stranded DNA viruses such as Maize streak virus (MSV) and various Gemini viruses. The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of MSV is thought to be involved in intra- and inter-cellular movement within the plant. Engineering of proteins having similarity to parvoviral coat proteins, especially to produce proteins that interfere with maturation of the virus particle, enables the production of plants having better resistance to natural plant single-stranded DNA viruses.

- Ref: Liu et al. (1997) J Gen Virol 78: 1265-70
Rohde et al. (1990) Virology 176: 648-51

Q. Pkinase_C

Plant serine/threonine protein kinases possessing this domain are expressed in all tissues and are known to undergo serine-specific autophosphorylation and specifically phosphorylate two ribosomal proteins, P14 and P16. During development, these proteins predominate during high metabolic activity in growing buds, root tips, leaf margins and germinating seeds. They are thought to be involved in the control of plant growth and development. In addition, two genes encoding proteins from this family have been described that help plant cells adapt during cold or high salt stresses. Consequently, engineering Pkinase C proteins provides a way to control general growth/development of the plant as well as a means to provide endogenous protection against environmental stresses.

- Ref: Zhang et al. (1994) J Biol Chem 269: 17586-92
Mizoguchi et al. (1995) FEBS Lett 358: 199-204

R. REV

The REV proteins act post-transcriptionally to relieve negative repression of GAG and ENV production in retroviruses such as Human Immunodeficiency Virus type I (HIV-1). Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutations at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant REV proteins enables control of transposition frequencies of corresponding transposable elements and provides a new tool for genetic engineering of plants.

- Ref: Sodroski et al. (1986) Nature 321: 412-7
Franchini et al. (1989) PNAS USA 86: 2433-7
Marquet et al. (1995) 77: 113-24

U. Signal

Many plant proteins in this family contain sequences similar to those found in both components of the prokaryotic family of signal transducers known as the two-component systems. This suggests that activation may require a transfer of a phosphate group between the transmitter domain and the receiver domain. One family member in Arabidopsis appears to be involved in ethylene (a plant hormone) signal transduction. Other proteins in this family appear to be involved in the regulation of gene transcription under conditions of environmental stress. Signal proteins can be exploited to affect plant growth and development and/or control plant responses to stress conditions such as cold, nutrient availability, etc.

Ref: Chang et al. (1993) Science 262: 539-44
Nagaya et al. (1993) Gene 131: 119-124
Gottfert et al. (1990) PNAS USA 87: 2680-4

V. vMSA

vMSA proteins are major surface antigens presenting on the envelope of various retroviruses. Surface antigens of retroviruses are often involved in tropism of the virus. Plants contain retrovirus-like viruses such as pararetroviruses and retrotransposons (i.e. transposons having long terminal repeats). Plant retrotransposons in particular have been used to create mutants at various loci, thereby permitting gene isolation, gene tagging and the like. Manipulation of plant vMSA proteins enables control of tropism of plant retroviruses that might be used for genetic engineering tools, thus enabling targeting of the virus to particular species and/or tissues of plants.

Ref: Okamoto et al. (1988) J Gen Virol 69: 2575-83
Grandbastien et al. (1989) Nature 337: 376-80
Wright and Voytas (1998) Genetics 149: 703-15

W. zf-CCCH

This family of proteins is defined by having two CX(8)CX(5)CX(3)H-type zinc finger domains. These proteins cover a broad range of functions. For example, the COP1 protein acts as a repressor of photomorphogenesis in darkness; light stimuli abolish this suppressive action. In addition, COP1 protein can function as a negative transcriptional regulator capable of direct interaction with components of the G-protein signaling pathway. As a second example, a zf-CCCH protein identified in Arabidopsis appears to be involved in the resistance to DNA damage induced by UV light and chemical DNA-damaging agents. Overexpression of this class of proteins permits production of plants that are better suited to adverse environments. Manipulation of expression of zf-CCCH proteins functioning as transcriptional regulators, such as COP1, enables manipulation of some signal transduction pathways.

Ref: Pang et al. (1993) Nucleic Acids Res 21: 1647-53

Deng et al. (1992) Cell 71: 791-801

X. zf-RanBP

Proteins falling within this category contain many X-X-F-G and X-F-X-F-G repeats, and may contain RANBP1-like or PPlase domains. Plant proteins having domains similar to these include PAS1 and GMSTI. PAS1 has been shown to have dramatic developmental affects that appear to be correlated with both cell division and cell wall elongation. GMSTI has high identity to the yeast STI stress-inducible gene and has been shown to be heat inducible. Proteins such as these may be useful for controlling growth and form of development.

Ref: Vittorioso et al. (1998) Mol Cell Biol 18: 3034-43

Hernandez Torres et al. (1995) 27: 1221-6

Y. Peptidase M48.

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are located in the membranes of the endoplasmic reticulum. They function in NH₂-terminal proteolytic processing, as shown for the yeast STE24 gene product. This gene is required for the correct processing of α -factor, a yeast pheromone. Family M48 peptidases

also appear to be required for some prenylation reactions, mediating COOH-terminal CAAX processing. Prenylation reactions are believed to be involved in the regulation of protein-protein and protein-membrane interactions. As an example, RAS GTPase activity is regulated in part by localization to the inner side of the plasma membrane upon prenylation.

In plants, proteins from this family could be involved in pollen-stigma interactions such as those mediating self-pollination vs. outcrossing, or could be members of several secondary metabolism pathways.

Ref: Fujimura-Kamada et al. (1997) J Cell Biol. 136: 271-85. Tam et al. (1998) J Cell Biol. 142: 635-49.

Z. DNA Pol Viral N

The DNA pol Viral N domain is located at the N-terminal region of DNA polymerase isolated from several retroviral viruses such as the Cauliflower Mosaic Virus. The domain motif has also been found in numerous other species from humans to cyanobacteria. In these organisms, this motif seems to be associated with two types of sequences; retrotransposons and mitochondrial genes. In the mitochondrial sequences this domain is potentially involved in the self-splicing conducted by group II introns. Various manipulations of this gene in plants allows control of the numerous retrotransposons endogenous to plant genomes or allows engineering of mitochondrial function, especially to increase efficiency of energy utilization by cells.

REF: Chapdelaine and Bonen (1991) Cell 65: 465-72
Ferat and Miche (1993) Nature 364: 358-61
Wilson et al. (1994) 368: 32-8
Cambareri et al. (1994) 242: 658-65
Gaardner et al. (1981) NAR 9: 2871-2888
Cummings et al. (1990) Curr Genet 17: 375-402
Hattori et al. (1986) Nature 321: 625-8

Aa. Calpain_inhib

This domain is found in calpastatin, an inhibitor protein specific for calpain. Calpain is a non-lysosomal calcium-dependent intracellular protease that appears to be involved in

the dynamic changes of the cytoskeleton, especially actin-related structures, during early *Drosophila* embryogenesis [1]. Calpastatins co-exist in cells with calpains and the subcellular distribution of calpastatin is thought to be important to calpain regulation [2]. In plants calpains and calpastatins could be involved in embryogenesis and non-embryogenic organ reiteration. Mutations occurring in calpain inhibitor repeat domains would produce developmental abnormalities such as abnormal leaf, root or flower development.

Refs

- 1 Emori Y and Saigo K (1994) J Biol Chem 269: 25137-42.
- 2 Mellgren RL, Lane RD, Mericle MT (1989) Biochim Biophys Acta 999: 71-77.

Ab. chorismate_bind

Chorismate binding domains are present in plant anthranilate synthase (AS) genes. AS genes catalyze the first step in the biosynthesis of tryptophan by converting chorismate and L-glutamine to anthranilate, pyruvate and L-glutamate. Some of these genes are involved in feedback inhibition by tryptophan [1] while some are feedback insensitive [2]. In Arabidopsis, two AS genes have overlapping, but different distributions. One of these AS genes is induced by wounding and bacterial pathogen infiltration [1]. Mutations in the chorismate binding domain would affect the production of tryptophan and could influence the plant's defense system. AS gene products can be used for *in vitro* synthesis of tryptophan and tryptophan derivatives.

Refs

- 1 Niyogi KK, Fink GR (1992) Plant Cell 4: 721-33.
- 2 Song HS, Brotherton JE, Gonzales RA, Willholm JM (1998) Plant Physiol 117:533-43.

Ac. late_protein_L2

Papillomaviruses are encapsulated double stranded DNA viruses. Plants are susceptible to infection by double stranded DNA viruses such as Cauliflower Mosaic virus (CaMV). The coat proteins in these plant viruses are critical to the virus life cycle within the plant. For example, the coat protein of CaMV is thought to be involved in intra- and inter-cellular movement within the plant [1]. Engineering of proteins having similarity to papillomavirus

coat proteins may enable the production of plants having better resistance to natural plant double stranded DNA viruses.

Refs

- 5 1 Thompson SR, Melcher U (1993) J Gen Virol 74: 1141-8.

Ad. Peptidase_M41

Proteins belonging to this peptidase family are metalloproteases that bind zinc as a cofactor and are integral membrane proteins. They seem to be involved in the degradation of carboxy-terminal-tagged cytoplasmic proteins. In plants, these proteins are located in the thylakoid membranes of the chloroplasts, their expression is light regulated and they are thought to be involved in degradation of soluble stromal proteins and turn-over of thylakoid proteins [1]. Manipulation of expression and structure of these proteins would have effects on the efficiency of photosynthesis and the development of chloroplasts.

Refs

- 1 Lindahl M, Tabak s, Cseke L, Pichersky E, Andersson B, Adam Z (1996) J Biol Chem 271: 29329-34.

Ae. UPF0051

There is some evidence that, in plants, proteins in this family are involved in ATP synthesis in chloroplasts [1, 2]. Mutations in these proteins or altering their expression would affect the efficiency of photosynthesis and energy production.

Refs

- 1 Kostrzewa M, Zetsche K (1992) J Mol Biol 227: 961-70.
2 Kostrzewa M, Zetsche K (1993) Plant Mol Biol 23: 67-76

Af. E7

Papillomaviruses are encapsulated double stranded DNA viruses. The Papillomavirus early protein 7 (E7) is known as a potent immortalizing and transforming agent. Transformation by E7 is thought to be mediated by the physical association of E7 with cellular proteins regulating entry into the cell cycle [1]. The result is entry into the cell cycle and suppression

of terminal differentiation in mammalian cells. Thus, engineering of proteins having similarity to papillomavirus E7 protein enables the production of plants having altered cellular proliferation characteristics and possibly altered morphology. For example, overexpression of E7-like proteins would be expected to result in proliferation of cells of the tissue in which the E7 protein is expressed, perhaps with suppression of differentiation events. Thus, for example, overexpression of E7-like proteins in meristem cells can result in taller plants and suppression of leafing and/or flowering.

Refs

1 Zwerschke W, Jansen-Durr P Adv Cancer Res 2000;78:1-29

Ag. Peptidase U7

This protein is known to be an integral membrane protein in the cyanobacterium Synechocystis where it functions to digest cleaved signal peptides [1]. This activity is necessary to maintain proper secretion of mature proteins across the membrane. In higher plants this protein may be present in the plastid or chloroplast membranes where it would function by enabling protein movement into and out of the chloroplasts. Mutations in this protein would be expected to affect the development of plastids, including chloroplasts, or alter the energy transfer system within the chloroplasts, thereby affecting growth and development.

Refs

1 Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirose M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) DNA Res 3:109-36.

Ah. 5'-3' Exonuclease

The 5'-3' exonuclease domain is one found in bacterial DNA polymerases I and in yeast DNA repair enzymes such as Exonuclease I. Yeast Exo I is involved in mitotic recombination and also includes a domain that interacts with the mismatch repair protein MSH2. The 5'-3' exonuclease domain is also present in XPG DNA repair enzymes in humans and in yeast RAD9 protein. Defects in XPG proteins result in Xeroderma Pigmentosum. Thus defects in 5'-3' exonuclease domain-containing proteins in plants are expected to lead to defects in DNA

repair and corresponding high spontaneous and inducible mutation rates. Consensus sequence:

IMKKKLLLVLDGSSLAFFALPPLTNSAGEPTNAVYGLKMLIKLIEQEQPTTHIAVV
 5 FDAKAKTFRHELYEGYKAGRAP
 TPDELREQIPLIKELLDALGIPLEVAGYEADDVIGTLAKLAEKEGYEVLIVTGDRDLL
 QLVSDHVTVIITKKGIAEFTL
 FTPEAVIEKYGLTPEQIIDYKALMGDSSDNIPGVKGIGEKTAAKLLOEYGSLEGYANL
 DKLKGKKLREKLLAHKEDAKL
 10 SRDLATIKTDVPLDLTLDDLRLPDPDRDALDLFDE

Ref:

Fiorentini P. et al. RT. Mol. Cell. Biol. 17:2764-2773(1997).

Tishkoff et al. Cancer Res. 0:0-0(1998).

15 Macinnes M.A. et al. Mol. Cell. Biol. 13:6393-6402(1993).

2025-09-24 10:00:00

AA. Activities of Polypeptides Comprising Signal Peptides

Polypeptides comprising signal peptides are a family of proteins that are typically targeted to (1) a particular organelle or intracellular compartment, (2) interact with a particular molecule or (3) for secretion outside of a host cell. Example of polypeptides comprising signal peptides include, without limitation, secreted proteins, soluble proteins, receptors, proteins retained in the ER, etc.

These proteins comprising signal peptides are useful to modulate ligand-receptor interactions, cell-to-cell communication, signal transduction, intracellular communication, and activities and/or chemical cascades that take part in an organism outside or within of any particular cell.

One class of such proteins are soluble proteins which are transported out of the cell. These proteins can act as ligands that bind to receptor to trigger signal transduction or to permit communication between cells.

Another class is receptor proteins which also comprise a retention domain that lodges the receptor protein in the membrane when the cell transports the receptor to the surface of the cell. Like the soluble ligands, receptors can also modulate signal transduction and communication between cells.

In addition the signal peptide itself can serve as a ligand for some receptors. An example is the interaction of the ER targeting signal peptide with the signal recognition particle (SRP). Here, the SRP binds to the signal peptide, halting translation, and the resulting SRP complex then binds to docking proteins located on the surface of the ER, prompting transfer of the protein into the ER.

A description of signal peptide residue composition is described below in Subsection IV.C.1.

III. Methods of Modulating Polypeptide Production

It is contemplated that polynucleotides of the invention can be incorporated into a host cell or in-vitro system to modulate polypeptide production. For instance, the SDFs prepared as described herein can be used to prepare expression cassettes useful in a number of techniques for suppressing or enhancing expression.

An example are polynucleotides comprising sequences to be transcribed, such as coding sequences, of the present invention can be inserted into nucleic acid constructs to modulate polypeptide production. Typically, such sequences to be transcribed are heterologous to at least one element of the nucleic acid construct to generate a chimeric gene or construct.

Another example of useful polynucleotides are nucleic acid molecules comprising regulatory sequences of the present invention. Chimeric genes or constructs can be generated when the regulatory sequences of the invention linked to heterologous sequences in a vector construct. Within the scope of invention are such chimeric gene and/or constructs.

Also within the scope of the invention are nucleic acid molecules, whereof at least a part or fragment of these DNA molecules are presented in TABLE 1 of the present application, and wherein the coding sequence is under the control of its own promoter and/or its own regulatory elements. Such molecules are useful for transforming the genome of a host cell or an organism regenerated from said host cell for modulating polypeptide production.

Additionally, a vector capable of producing the oligonucleotide can be inserted into the host cell to deliver the oligonucleotide.

More detailed description of components to be included in vector constructs are described both above and below.

Whether the chimeric vectors or native nucleic acids are utilized, such polynucleotides can be incorporated into a host cell to modulate polypeptide production. Native genes and/or nucleic acid molecules can be effective when exogenous to the host cell.

Methods of modulating polypeptide expression includes, without limitation:

Suppression methods, such as

Antisense

Ribozymes

Co-suppression

Insertion of Sequences into the Gene to be Modulated

Regulatory Sequence Modulation.

as well as Methods for Enhancing Production, such as

Insertion of Exogenous Sequences; and

Regulatory Sequence Modulation.

III.A. Suppression

Expression cassettes of the invention can be used to suppress expression of endogenous genes which comprise the SDF sequence. Inhibiting expression can be useful, for instance, to tailor the ripening characteristics of a fruit (Oeller et al., *Science* 254:437 (1991)) or to influence seed size_(WO98/07842) or to provoke cell ablation (Mariani et al., *Nature* 357: 384-387 (1992)).

As described above, a number of methods can be used to inhibit gene expression in plants, such as antisense, ribozyme, introduction of exogenous genes into a host cell, insertion of a polynucleotide sequence into the coding sequence and/or the promoter of the endogenous gene of interest, and the like.

III.A.1. Antisense

An expression cassette as described above can be transformed into host cell or plant to produce an antisense strand of RNA. For plant cells, antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, *see*, e.g., Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

III.A.2. Ribozymes

Similarly, ribozyme constructs can be transformed into a plant to cleave mRNA and down-regulate translation.

III.A.3. Co-Suppression

Another method of suppression is by introducing an exogenous copy of the gene to be suppressed. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to prevent the accumulation of mRNA. A detailed description of this method is described above.

III.A.4. Insertion of Sequences into the Gene to be Modulated

Yet another means of suppressing gene expression is to insert a polynucleotide into the gene of interest to disrupt transcription or translation of the gene.

Homologous recombination could be used to target a polynucleotide insert to a gene using the Cre-Lox system (A.C. Vergunst et al., *Nucleic Acids Res.* **26**:2729 (1998), A.C. Vergunst et al., *Plant Mol. Biol.* **38**:393 (1998), H. Albert et al., *Plant J.* **7**:649 (1995)).

In addition, random insertion of polynucleotides into a host cell genome can also be used to disrupt the gene of interest. Azpiroz-Lechan et al., *Trends in Genetics* **13**:152 (1997). In this method, screening for clones from a library containing random insertions is preferred for identifying those that have polynucleotides inserted into the gene of interest. Such screening can be performed using probes and/or primers described above based on sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto. The screening can also be performed by selecting clones or any transgenic plants having a desired phenotype.

III.A.5. Regulatory Sequence Modulation

The SDFs described in Table 1, and fragments thereof are examples of nucleotides of the invention that contain regulatory sequences that can be used to suppress or inactivate transcription and/or translation from a gene of interest as discussed in I.C.5.

III.A.6. Genes Comprising Dominant-Negative Mutations

When suppression of production of the endogenous, native protein is desired it is often helpful to express a gene comprising a dominant negative mutation. Production of protein variants produced from genes comprising dominant negative mutations is a useful tool for research. Genes comprising dominant negative mutations can produce a variant polypeptide which is capable of competing with the native polypeptide, but which does not produce the native result. Consequently, over expression of genes comprising these mutations can titrate out an undesired activity of the native protein. For example, The product from a gene comprising a dominant negative mutation of a receptor can be used to constitutively activate or suppress a signal transduction cascade, allowing examination of the phenotype and thus the trait(s) controlled by that receptor and pathway. Alternatively, the protein arising from the gene comprising a dominant-negative mutation can be an inactive enzyme still capable

of binding to the same substrate as the native protein and therefore competes with such native protein.

Products from genes comprising dominant-negative mutations can also act upon the native protein itself to prevent activity. For example, the native protein may be active only as a homo-multimer or as one subunit of a hetero-multimer. Incorporation of an inactive subunit into the multimer with native subunit(s) can inhibit activity.

Thus, gene function can be modulated in host cells of interest by insertion into these cells vector constructs comprising a gene comprising a dominant-negative mutation.

III.B. Enhanced Expression

Enhanced expression of a gene of interest in a host cell can be accomplished by either (1) insertion of an exogenous gene; or (2) promoter modulation.

III.B.1. Insertion of an Exogenous Gene

Insertion of an expression construct encoding an exogenous gene can boost the number of gene copies expressed in a host cell.

Such expression constructs can comprise genes that either encode the native protein that is of interest or that encode a variant that exhibits enhanced activity as compared to the native protein. Such genes encoding proteins of interest can be constructed from the sequences from TABLE 1, fragments thereof, and substantially similar sequence thereto.

Such an exogenous gene can include either a constitutive promoter permitting expression in any cell in a host organism or a promoter that directs transcription only in particular cells or times during a host cell life cycle or in response to environmental stimuli.

III.B.2. Regulatory Sequence Modulation

The SDFs of Table 1, and fragments thereof, contain regulatory sequences that can be used to enhance expression of a gene of interest. For example, some of these sequences contain useful enhancer elements. In some cases, duplication of enhancer elements or insertion of exogenous enhancer elements will increase expression of a desired gene from a particular promoter. As other examples, all II promoters require binding of a regulatory protein to be activated, while some promoters may need a protein that signals a promoter binding protein to expose a polymerase binding site. In either case, over-production of such proteins can be used to enhance expression of a gene of interest by increasing the activation time of the promoter.

Such regulatory proteins are encoded by some of the sequences in TABLE 1, fragments thereof, and substantially similar sequences thereto.

Coding sequences for these proteins can be constructed as described above.

IV. Gene Constructs and Vector Construction

To use isolated SDFs of the present invention or a combination of them or parts and/or mutants and/or fusions of said SDFs in the above techniques, recombinant DNA vectors which comprise said SDFs and are suitable for transformation of cells, such as plant cells, are usually prepared. The SDF construct can be made using standard recombinant DNA techniques (Sambrook et al. 1989) and can be introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation (e.g., particle gun bombardment) as referenced below.

The vector backbone can be any of those typical in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

- (a) **BAC:** Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-8797 (1992); Hamilton et al., Proc. Natl. Acad. Sci. USA 93: 9975-9979 (1996);
- (b) **YAC:** Burke et al., Science 236:806-812 (1987);.
- (c) **PAC:** Sternberg N. et al., Proc Natl Acad Sci U S A. Jan;87(1):103-7 (1990);
- (d) **Bacteria-Yeast Shuttle Vectors:** Bradshaw et al., Nucl Acids Res 23: 4850-4856 (1995);
- (e) **Lambda Phage Vectors:** Replacement Vector, e.g., Frischauf et al., J. Mol Biol 170: 827-842 (1983); or Insertion vector, e.g., Huynh et al., In: Glover NM (ed) DNA Cloning: A practical Approach, Vol.1 Oxford: IRL Press (1985);
- (f) **T-DNA gene fusion vectors:** Walden et al., Mol Cell Biol 1: 175-194 (1990); and
- (g) **Plasmid vectors:** Sambrook et al., infra.

Typically, a vector will comprise the exogenous gene, which in its turn comprises an SDF of the present invention to be introduced into the genome of a host cell, and which gene may be an antisense construct, a ribozyme construct chimera, or a coding sequence with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs,

and 3' end termination sequences. Vectors of the invention can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc.

A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for over-expression, a plant promoter fragment may be employed that will direct transcription of the gene in all tissues of a regenerated plant. Alternatively, the plant promoter may direct transcription of an SDF of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters).

If proper polypeptide production is desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from genes or SDF of the invention may comprise a marker gene that confers a selectable phenotype on plant cells. The vector can include promoter and coding sequence, for instance. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or phosphinotricin.

IV.A. Coding Sequences

Generally, the sequence in the transformation vector and to be introduced into the genome of the host cell does not need to be absolutely identical to an SDF of the present invention. Also, it is not necessary for it to be full length, relative to either the primary transcription product or fully processed mRNA. Furthermore, the introduced sequence need not have the same intron or exon pattern as a native gene. Also, heterologous non-coding segments can be incorporated into the coding sequence without changing the desired amino acid sequence of the polypeptide to be produced.

IV.B. Promoters

As explained above, introducing an exogenous SDF from the same species or an orthologous SDF from another species can modulate the expression of a native gene

corresponding to that SDF of interest. Such an SDF construct can be under the control of either a constitutive promoter or a highly regulated inducible promoter (e.g., a copper inducible promoter). The promoter of interest can initially be either endogenous or heterologous to the species in question. When re-introduced into the genome of said species, such promoter becomes exogenous to said species. Over-expression of an SDF transgene can lead to co-suppression of the homologous endogeneous sequence thereby creating some alterations in the phenotypes of the transformed species as demonstrated by similar analysis of the chalcone synthase gene (Napoli et al., *Plant Cell* 2:279 (1990) and van der Krol et al., *Plant Cell* 2:291 (1990)). If an SDF is found to encode a protein with desirable characteristics, its over-production can be controlled so that its accumulation can be manipulated in an organ- or tissue-specific manner utilizing a promoter having such specificity.

Likewise, if the promoter of an SDF (or an SDF that includes a promoter) is found to be tissue-specific or developmentally regulated, such a promoter can be utilized to drive or facilitate the transcription of a specific gene of interest (e.g., seed storage protein or root-specific protein). Thus, the level of accumulation of a particular protein can be manipulated or its spatial localization in an organ- or tissue- specific manner can be altered.

IV. C. Signal Peptides

SDFs of the present invention containing signal peptides are indicated in Table 1. In some cases it may be desirable for the protein encoded by an introduced exogenous or orthologous SDF to be targeted (1) to a particular organelle intracellular compartment, (2) to interact with a particular molecule such as a membrane molecule or (3) for secretion outside of the cell harboring the introduced SDF. This will be accomplished using a signal peptide.

Signal peptides direct protein targeting, are involved in ligand-receptor interactions and act in cell to cell communication. Many proteins, especially soluble proteins, contain a signal peptide that targets the protein to one of several different intracellular compartments. In plants, these compartments include, but are not limited to, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, protein storage vesicles (PSV) and, in general, membranes. Some signal peptide sequences are conserved, such as the Asn-Pro-Ile-Arg amino acid motif found in the N-terminal propeptide signal that targets proteins to the vacuole (Marty (1999) *The Plant Cell* 11: 587-599). Other signal peptides do not have a consensus sequence *per se*, but are largely composed of

hydrophobic amino acids, such as those signal peptides targeting proteins to the ER (Vitale and Denecke (1999) *The Plant Cell* 11: 615-628). Still others do not appear to contain either a consensus sequence or an identified common secondary sequence, for instance the chloroplast stromal targeting signal peptides (Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). Furthermore, some targeting peptides are bipartite, directing proteins first to an organelle and then to a membrane within the organelle (e.g. within the thylakoid lumen of the chloroplast; see Keegstra and Cline (1999) *The Plant Cell* 11: 557-570). In addition to the diversity in sequence and secondary structure, placement of the signal peptide is also varied. Proteins destined for the vacuole, for example, have targeting signal peptides found at the N-terminus, at the C-terminus and at a surface location in mature, folded proteins. Signal peptides also serve as ligands for some receptors.

These characteristics of signal proteins can be used to more tightly control the phenotypic expression of introduced SDFs. In particular, associating the appropriate signal sequence with a specific SDF can allow sequestering of the protein in specific organelles (plastids, as an example), secretion outside of the cell, targeting interaction with particular receptors, etc. Hence, the inclusion of signal proteins in constructs involving the SDFs of the invention increases the range of manipulation of SDF phenotypic expression. The nucleotide sequence of the signal peptide can be isolated from characterized genes using common molecular biological techniques or can be synthesized in vitro.

In addition, the native signal peptide sequences, both amino acid and nucleotide, described in Table 1 can be used to modulate polypeptide transport. Further variants of the native signal peptides described in Table 1 are contemplated. Insertions, deletions, or substitutions can be made. Such variants will retain at least one of the functions of the native signal peptide as well as exhibiting some degree of sequence identity to the native sequence.

Also, fragments of the signal peptides of the invention are useful and can be fused with other signal peptides of interest to modulate transport of a polypeptide.

V. Transformation Techniques

A wide range of techniques for inserting exogenous polynucleotides are known for a number of host cells, including, without limitation, bacterial, yeast, mammalian, insect and plant cells.

Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g. Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, *Euphytica*, v. 85, n.1-3:13-27, (1995).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., *Mol. Biotechnol.* 8:199 (1997); Hamilton, *Gene* 200:107 (1997)); Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983).

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, *CM., Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, *AP., Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in "Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73,

CRC Press, Boca Raton, 1988. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* **38**:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* **58**:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* **32**:1 (1994)). The nucleic acids of the invention can be used to confer desired traits on essentially any plant.

Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and, *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The particular sequences of SDFs identified are provided in the attached TABLE 1. One of ordinary skill in the art, having this data, can obtain cloned DNA fragments, synthetic DNA fragments or polypeptides constituting desired sequences by recombinant methodology known in the art or described herein.

EXAMPLES

The invention is illustrated by way of the following examples. The invention is not limited by these examples as the scope of the invention is defined solely by the claims following.

EXAMPLE 1: cDNA PREPARATION

A number of the nucleotide sequences disclosed in TABLE 1 herein as representative of the SDFs of the invention can be obtained by sequencing genomic DNA (gDNA) and/or cDNA from corn plants grown from HYBRID SEED # 35A19, purchased from Pioneer Hi-Bred International, Inc., Supply Management, P.O. Box 256, Johnston, Iowa 50131-0256.

A number of the nucleotide sequences disclosed in TABLE 1 herein as representative of the SDFs of the invention can also be obtained by sequencing genomic DNA from *Arabidopsis thaliana*, Wassilewskija ecotype or by sequencing cDNA obtained from mRNA from such plants as described below. This is a true breeding strain. Seeds of the plant are available from the Arabidopsis Biological Resource Center at the Ohio State University, under the accession number CS2360. Seeds of this plant were deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection, Manassas, VA on August 31, 1999, and were assigned ATCC No. PTA-595.

Other methods for cloning full-length cDNA are described, for example, by Seki et al., *Plant Journal* 15:707-720 (1998) "High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated Cap trapper"; Maruyama et al., *Gene* 138:171 (1994) "Oligo-capping a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides"; and WO 96/34981.

Tissues were, or each organ was, individually pulverized and frozen in liquid nitrogen. Next, the samples were homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed. Then the sample was applied to a 2M sucrose cushion to isolate polysomes. The RNA was isolated by treatment with detergents and proteinase K followed by ethanol precipitation and centrifugation. The polysomal RNA from the different tissues was pooled according to the following mass ratios: 15/15/1 for male inflorescences, female inflorescences and root, respectively. The pooled material was then used for cDNA synthesis by the methods described below.

Starting material for cDNA synthesis for the exemplary corn cDNA clones with sequences presented in TABLE 1 was poly(A)-containing polysomal mRNAs from inflorescences and root tissues of corn plants grown from HYBRID SEED # 35A19. Male inflorescences and female (pre-and post-fertilization) inflorescences were isolated at various stages of development. Selection for poly(A) containing polysomal RNA was done using oligo d(T) cellulose columns, as described by Cox and Goldberg, "Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The quality and the integrity of the polyA+ RNAs were evaluated.

Starting material for cDNA synthesis for the exemplary *Arabidopsis* cDNA clones with sequences presented in TABLE 1 was polysomal RNA isolated from the top-most inflorescence tissues of *Arabidopsis thaliana* Wassilewskija (Ws.) and from roots of *Arabidopsis thaliana* Landsberg erecta (L. er.), also obtained from the Arabidopsis Biological Resource Center. Nine parts inflorescence to every part root was used, as measured by wet mass. Tissue was pulverized and exposed to liquid nitrogen. Next, the sample was homogenized in the presence of detergents and then centrifuged. The debris and nuclei were removed from the sample and more detergents were added to the sample. The sample was centrifuged and the debris was removed and the sample was applied to a 2M sucrose cushion to isolate polysomal RNA. Cox et al., "Plant Molecular Biology: A Practical Approach", pp. 1-35, Shaw ed., c. 1988 by IRL, Oxford. The polysomal RNA was used for cDNA synthesis by the methods described below. Polysomal mRNA was then isolated as described above for corn cDNA. The quality of the RNA was assessed electrophoretically.

Following preparation of the mRNAs from various tissues as described above, selection of mRNA with intact 5' ends and specific attachment of an oligonucleotide tag to the 5' end of such mRNA was performed using either a chemical or enzymatic approach. Both techniques take advantage of the presence of the "cap" structure, which characterizes the 5' end of most intact mRNAs and which comprises a guanosine generally methylated once, at the 7 position.

The chemical modification approach involves the optional elimination of the 2', 3'-cis diol of the 3' terminal ribose, the oxidation of the 2', 3'-cis diol of the ribose linked to the cap of the 5' ends of the mRNAs into a dialdehyde, and the coupling of the such obtained dialdehyde to a derivatized oligonucleotide tag. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in International Application No. WO96/34981 published November 7, 1996.

The enzymatic approach for ligating the oligonucleotide tag to the intact 5' ends of mRNAs involves the removal of the phosphate groups present on the 5' ends of uncapped incomplete mRNAs, the subsequent decapping of mRNAs having intact 5' ends and the ligation of the phosphate present at the 5' end of the decapped mRNA to an oligonucleotide tag. Further detail regarding the enzymatic approaches for obtaining mRNAs having intact 5' ends are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato et al., *Gene* 150:243-250 (1994).

In both the chemical and the enzymatic approach, the oligonucleotide tag has a restriction enzyme site (e.g. an EcoRI site) therein to facilitate later cloning procedures. Following attachment of the oligonucleotide tag to the mRNA, the integrity of the mRNA is examined by performing a Northern blot using a probe complementary to the oligonucleotide tag.

For the mRNAs joined to oligonucleotide tags using either the chemical or the enzymatic method, first strand cDNA synthesis is performed using an oligo-dT primer with reverse transcriptase. This oligo-dT primer can contain an internal tag of at least 4 nucleotides, which can be different from one mRNA preparation to another. Methylated dCTP is used for cDNA first strand synthesis to protect the internal EcoRI sites from digestion during subsequent steps. The first strand cDNA is precipitated using isopropanol after removal of RNA by alkaline hydrolysis to eliminate residual primers.

Second strand cDNA synthesis is conducted using a DNA polymerase, such as Klenow fragment and a primer corresponding to the 5' end of the ligated oligonucleotide. The primer is typically 20-25 bases in length. Methylated dCTP is used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following second strand synthesis, the full-length cDNAs are cloned into a phagemid vector, such as pBlueScript™ (Stratagene). The ends of the full-length cDNAs are blunted with T4 DNA polymerase (Biolabs) and the cDNA is digested with EcoRI. Since methylated dCTP is used during cDNA synthesis, the EcoRI site present in the tag is the only hemi-methylated site; hence the only site susceptible to EcoRI digestion. In some instances, to facilitate subcloning, an Hind III adapter is added to the 3' end of full-length cDNAs.

The full-length cDNAs are then size fractionated using either exclusion chromatography (AcA, Biosepra) or electrophoretic separation which yields 3 to 6 different fractions. The full-length cDNAs are then directionally cloned either into pBlueScript™ using either the EcoRI and SmaI restriction sites or, when the Hind III adapter is present in the full-length cDNAs, the EcoRI and Hind III restriction sites. The ligation mixture is transformed, preferably by electroporation, into bacteria, which are then propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached to full-length cDNAs are selected as follows.

The plasmid cDNA libraries made as described above are purified (e.g. by a column available from Qiagen). A positive selection of the tagged clones is performed as follows. Briefly, in this selection procedure, the plasmid DNA is converted to single stranded DNA using

phage F1 gene II endonuclease in combination with an exonuclease (Chang et al., *Gene* 127:95 (1993)) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA is then purified using paramagnetic beads as described by Fry et al., *Biotechniques* 13: 124 (1992). Here the single stranded DNA is hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide tag. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide are selected by incubation with streptavidin coated magnetic beads followed by magnetic capture. After capture of the positive clones, the plasmid DNA is released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as ThermoSequenase™ (obtained from Amersham Pharmacia Biotech). Alternatively, protocols such as the Gene Trapper™ kit (Gibco BRL) can be used. The double stranded DNA is then transformed, preferably by electroporation, into bacteria. The percentage of positive clones having the 5' tag oligonucleotide is typically estimated to be between 90 and 98% from dot blot analysis.

Following transformation, the libraries are ordered in microtiter plates and sequenced. The *Arabidopsis* library was deposited at the American Type Culture Collection on January 7, 2000 as "*E. coli* liba 010600" under the accession number PTA-1161.

EXAMPLE 2: SOUTHERN HYBRIDIZATIONS

The SDFs of the invention can be used in Southern hybridizations as described above. The following describes extraction of DNA from nuclei of plant cells, digestion of the nuclear DNA and separation by length, transfer of the separated fragments to membranes, preparation of probes for hybridization, hybridization and detection of the hybridized probe.

The procedures described herein can be used to isolate related polynucleotides or for diagnostic purposes. Moderate stringency hybridization conditions, as defined above, are described in the present example. These conditions result in detection of hybridization between sequences having at least 70% sequence identity. As described above, the hybridization and wash conditions can be changed to reflect the desired percentage of sequence identity between probe and target sequences that can be detected.

In the following procedure, a probe for hybridization is produced from two PCR reactions using two primers from genomic sequence of *Arabidopsis thaliana*. As described above, the particular template for generating the probe can be any desired template.

The first PCR product is assessed to validate the size of the primer to assure it is of the expected size. Then the product of the first PCR is used as a template, with the same pair

850

of primers used in the first PCR, in a second PCR that produces a labeled product used as the probe.

Fragments detected by hybridization, or other bands of interest, can be isolated from gels used to separate genomic DNA fragments by known methods for further purification and/or characterization.

Buffers for nuclear DNA extraction

1. 10X HB

	1000 ml	
40 mM spermidine	10.2 g	Spermine (Sigma S-2876) and spermidine (Sigma S-2501)
10 mM spermine	3.5 g	Stabilize chromatin and the nuclear membrane
0.1 M EDTA (disodium)	37.2 g	EDTA inhibits nuclease
0.1 M Tris	12.1 g	Buffer
0.8 M KCl	59.6 g	Adjusts ionic strength for stability of nuclei

Adjust pH to 9.5 with 10 N NaOH. It appears that there is a nuclease present in leaves. Use of pH 9.5 appears to inactivate this nuclease.

2. 2 M sucrose (684 g per 1000 ml)

Heat about half the final volume of water to about 50°C. Add the sucrose slowly then bring the mixture to close to final volume; stir constantly until it has dissolved. Bring the solution to volume.

3. Sarkosyl solution (lyses nuclear membranes)

	851	
N-lauroyl sarcosine (Sarkosyl)		20.0 g
0.1 M Tris		12.1 g
0.04 M EDTA (Disodium)	14.9 g	

Adjust the pH to 9.5 after all the components are dissolved and bring up to the proper volume.

4. 20% Triton X-100
80 ml Triton X-100
320 ml 1xHB (w/o β -ME and PMSF)
Prepare in advance; Triton takes some time to dissolve

A. Procedure

1. Prepare 1X "H" buffer (keep ice-cold during use)

	<u>1000 ml</u>	
10X HB		100 ml
2 M sucrose	250 ml	a non-ionic osmoticum
Water	634 ml	

Added just before use:

100 mM PMSF*	10 ml	a protease inhibitor; protects nuclear membrane proteins
β -mercaptoethanol	1 ml	inactivates nuclease by reducing disulfide bonds

*100 mM PMSF

(phenyl methyl sulfonyl fluoride, Sigma P-7626)

(add 0.0875 g to 5 ml 100% ethanol)

2. Homogenize the tissue in a blender (use 300-400 ml of 1xHB per blender). Be sure that you use 5-10 ml of HB buffer per gram of tissue. Blenders generate heat so be

sure to keep the homogenate cold. It is necessary to put the blenders in ice periodically.

3. Add the 20% Triton X-100 (25 ml per liter of homogenate) and gently stir on ice for 20 min. This lyses plastid, but not nuclear, membranes.

4. Filter the tissue suspension through several nylon filters into an ice-cold beaker. The first filtration is through a 250-micron membrane; the second is through an 85-micron membrane; the third is through a 50-micron membrane; and the fourth is through a 20-micron membrane. Use a large funnel to hold the filters. Filtration can be sped up by gently squeezing the liquid through the filters.

5. Centrifuge the filtrate at 1200 x g for 20 min. at 4°C to pellet the nuclei.

6. Discard the dark green supernatant. The pellet will have several layers to it. One is starch; it is white and gritty. The nuclei are gray and soft. In the early steps, there may be a dark green and somewhat viscous layer of chloroplasts.

Wash the pellets in about 25 ml cold H buffer (with Triton X-100) and resuspend by swirling gently and pipetting. After the pellets are resuspended.

Pellet the nuclei again at 1200 - 1300 x g. Discard the supernatant.

Repeat the wash 3-4 times until the supernatant has changed from a dark green to a pale green. This usually happens after 3 or 4 resuspensions. At this point, the pellet is typically grayish white and very slippery. The Triton X-100 in these repeated steps helps to destroy the chloroplasts and mitochondria that contaminate the prep.

Resuspend the nuclei for a final time in a total of 15 ml of H buffer and transfer the suspension to a sterile 125 ml Erlenmeyer flask.

7. Add 15 ml, dropwise, cold 2% Sarkosyl, 0.1 M Tris, 0.04 M EDTA solution (pH 9.5) while swirling gently. This lyses the nuclei. The solution will become very viscous.

8. Add 30 grams of CsCl and gently swirl at room temperature until the CsCl is in solution. The mixture will be gray, white and viscous.
9. Centrifuge the solution at 11,400 x g at 4°C for at least 30 min. The longer this spin is, the firmer the protein pellicle.
10. The result is typically a clear green supernatant over a white pellet, and (perhaps) under a protein pellicle. Carefully remove the solution under the protein pellicle and above the pellet. Determine the density of the solution by weighing 1 ml of solution and add CsCl if necessary to bring to 1.57 g/ml. The solution contains dissolved solids (sucrose etc) and the refractive index alone will not be an accurate guide to CsCl concentration.
11. Add 20 µl of 10 mg/ml EtBr per ml of solution.
12. Centrifuge at 184,000 x g for 16 to 20 hours in a fixed-angle rotor.
13. Remove the dark red supernatant that is at the top of the tube with a plastic transfer pipette and discard. Carefully remove the DNA band with another transfer pipette. The DNA band is usually visible in room light; otherwise, use a long wave UV light to locate the band.
14. Extract the ethidium bromide with isopropanol saturated with water and salt. Once the solution is clear, extract at least two more times to ensure that all of the EtBr is gone. Be very gentle, as it is very easy to shear the DNA at this step. This extraction may take a while because the DNA solution tends to be very viscous. If the solution is too viscous, dilute it with TE.
15. Dialyze the DNA for at least two days against several changes (at least three times) of TE (10 mM Tris, 1mM EDTA, pH 8) to remove the cesium chloride.

16. Remove the dialyzed DNA from the tubing. If the dialyzed DNA solution contains a lot of debris, centrifuge the DNA solution at least at 2500 x g for 10 min. and carefully transfer the clear supernatant to a new tube. Read the A₂₆₀ concentration of the DNA.
17. Assess the quality of the DNA by agarose gel electrophoresis (1% agarose gel) of the DNA. Load 50 ng and 100 ng (based on the OD reading) and compare it with known and good quality DNA. Undigested lambda DNA and a lambda-HindIII-digested DNA are good molecular weight makers.

Protocol for Digestion of Genomic DNA

Protocol:

1. The relative amounts of DNA for different crop plants that provide approximately a balanced number of genome equivalent is given in Table 3. Note that due to the size of the wheat genome, wheat DNA will be underrepresented. Lambda DNA provides a useful control for complete digestion.
2. Precipitate the DNA by adding 3 volumes of 100% ethanol. Incubate at -20°C for at least two hours. Yeast DNA can be purchased and made up at the necessary concentration, therefore no precipitation is necessary for yeast DNA.
3. Centrifuge the solution at 11,400 x g for 20 min. Decant the ethanol carefully (be careful not to disturb the pellet). Be sure that the residual ethanol is completely removed either by vacuum desiccation or by carefully wiping the sides of the tubes with a clean tissue.
4. Resuspend the pellet in an appropriate volume of water. Be sure the pellet is fully resuspended before proceeding to the next step. This may take about 30 min.
5. Add the appropriate volume of 10X reaction buffer provided by the manufacturer of the restriction enzyme to the resuspended DNA followed by the appropriate volume of enzymes. Be sure to mix it properly by slowly swirling the tubes.

6. Set-up the lambda digestion-control for each DNA that you are digesting.
7. Incubate both the experimental and lambda digests overnight at 37°C. Spin down condensation in a microfuge before proceeding.
8. After digestion, add 2 µl of loading dye (typically 0.25% bromophenol blue, 0.25% xylene cyanol in 15% Ficoll or 30% glycerol) to the lambda-control digests and load in 1% TPE-agarose gel (TPE is 90 mM Tris-phosphate, 2 mM EDTA, pH 8). If the lambda DNA in the lambda control digests are completely digested, proceed with the precipitation of the genomic DNA in the digests.
9. Precipitate the digested DNA by adding 3 volumes of 100% ethanol and incubating in -20°C for at least 2 hours (preferably overnight).

EXCEPTION: *Arabidopsis* and yeast DNA are digested in an appropriate volume; they don't have to be precipitated.

10. Resuspend the DNA in an appropriate volume of TE (e.g., 22 µl x 50 blots = 1100 µl) and an appropriate volume of 10X loading dye (e.g., 2.4 µl x 50 blots = 120 µl). Be careful in pipetting the loading dye - it is viscous. Be sure you are pipetting the correct volume.

Table 3

Some guide points in digesting genomic DNA.

Species	Genome Size	Size Relative to <i>Arabidopsis</i>	Genome Equivalent to 2 µg <i>Arabidopsis</i> DNA	Amount of DNA per blot
<i>Arabidopsis</i>	120 Mb	1X	1X	2 µg
Brassica	1,100 Mb	9.2X	0.54X	10 µg
Corn	2,800 Mb	23.3X	0.43X	20 µg

856

Cotton	2,300 Mb	19.2X	0.52X	20 µg
Oat	11,300 Mb	94X	0.11X	20 µg
Rice	400 Mb	3.3X	0.75X	5 µg
Soybean	1,100 Mb	9.2X	0.54X	10 µg
Sugarbeet	758 Mb	6.3X	0.8X	10 µg
Sweetclover	1,100 Mb	9.2X	0.54X	10 µg
Wheat	16,000 Mb	133X	0.08X	20 µg
Yeast	15 Mb	0.12X	1X	0.25 µg

Protocol for Southern Blot Analysis

The digested DNA samples are electrophoresed in 1% agarose gels in 1x TPE buffer. Low voltage; overnight separations are preferred. The gels are stained with EtBr and photographed.

- For blotting the gels, first incubate the gel in 0.25 N HCl (with gentle shaking) for about 15 min.
- Then briefly rinse with water. The DNA is denatured by 2 incubations. Incubate (with shaking) in 0.5 M NaOH in 1.5 M NaCl for 15 min.
- The gel is then briefly rinsed in water and neutralized by incubating twice (with shaking) in 1.5 M Tris pH 7.5 in 1.5 M NaCl for 15 min.
- A nylon membrane is prepared by soaking it in water for at least 5 min, then in 6X SSC for at least 15 min. before use. (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate per liter, adjusted to pH 7.0.)
- The nylon membrane is placed on top of the gel and all bubbles in between are removed. The DNA is blotted from the gel to the membrane using an absorbent medium, such as paper toweling and 6x SCC buffer. After the transfer, the membrane may be lightly brushed with a gloved hand to remove any agarose sticking to the surface.

6. The DNA is then fixed to the membrane by UV crosslinking and baking at 80°C. The membrane is stored at 4°C until use.

B. Protocol for PCR Amplification of Genomic Fragments in Arabidopsis

Amplification procedures:

- 5 1. Mix the following in a 0.20 ml PCR tube or 96-well PCR plate:

Volume	Stock	Final Amount or Conc.
0.5 µl	~ 10 ng/µl genomic DNA ¹	5 ng
2.5 µl	10X PCR buffer	20 mM Tris, 50 mM KCl
0.75 µl	50 mM MgCl ₂	1.5 mM
1 µl	10 pmol/µl Primer 1 (Forward)	10 pmol
1 µl	10 pmol/µl Primer 2 (Reverse)	10 pmol
0.5 µl	5 mM dNTPs	0.1 mM
0.1 µl	5 units/µl Platinum Taq™ (Life Technologies, Gaithersburg, MD) DNA Polymerase	1 units
(to 25 µl)	Water	

2. The template DNA is amplified using a Perkin Elmer 9700 PCR machine:

¹ Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

- 1) 94°C for 10 min. followed by

2)	3)	4)
5 cycles:	5 cycles:	25 cycles:
94 °C - 30 sec	94 °C - 30 sec	94 °C - 30 sec
62 °C - 30 sec	58 °C - 30 sec	53 °C - 30 sec
72 °C - 3 min	72 °C - 3 min	72 °C - 3 min

- 5) 72°C for 7 min. Then the reactions are stopped by chilling to 4°C.

The procedure can be adapted to a multi-well format if necessary.

Quantification and Dilution of PCR Products:

- The product of the PCR is analyzed by electrophoresis in a 1% agarose gel. A linearized plasmid DNA can be used as a quantification standard (usually at 50, 100, 200, and 400 ng). These will be used as references to approximate the amount of PCR products. HindIII-digested Lambda DNA is useful as a molecular weight marker. The gel can be run fairly quickly; e.g., at 100 volts. The standard gel is examined to determine that the size of the PCR products is consistent with the expected size and if there are significant extra bands or smeary products in the PCR reactions.
- The amounts of PCR products can be estimated on the basis of the plasmid standard.
- For the small number of reactions that produce extraneous bands, a small amount of DNA from bands with the correct size can be isolated by dipping a sterile 10- μ l tip into the band while viewing through a UV Transilluminator. The small amount of agarose gel (with the DNA fragment) is used in the labeling reaction.

C. Protocol for PCR-DIG-Labeling of DNA

Solutions:

Reagents in PCR reactions (diluted PCR products, 10X PCR Buffer, 50 mM MgCl₂, 5 U/μl Platinum Taq Polymerase, and the primers)

10X dNTP + DIG-11-dUTP [**1:5**]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.65 mM dTTP, 0.35 mM DIG-11-dUTP)

5 10X dNTP + DIG-11-dUTP [**1:10**]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.81 mM dTTP, 0.19 mM DIG-11-dUTP)

10X dNTP + DIG-11-dUTP [**1:15**]: (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.875 mM dTTP, 0.125 mM DIG-11-dUTP)

TE buffer (10 mM Tris, 1 mM EDTA, pH 8)

10 Maleate buffer: In 700 ml of deionized distilled water, dissolve 11.61 g maleic acid and 8.77 g NaCl. Add NaOH to adjust the pH to 7.5. Bring the volume to 1 L. Stir for 15 min, and sterilize.

15 10% blocking solution: In 80 ml deionized distilled water, dissolve 1.16g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, Cat. no. 1096176). Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

1% blocking solution: Dilute the 10% stock to 1% using the maleate buffer.

20 Buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM MgCl₂, pH9.5). Prepared from autoclaved solutions of 1M Tris pH 9.5, 5 M NaCl, and 1 M MgCl₂ in autoclaved distilled water.

Procedure:

1. PCR reactions are performed in 25 µl volumes containing:

PCR buffer	1X
MgCl ₂	1.5 mM
10X dNTP + DIG-11-dUTP	1X (please see the note below)
Platinum Taq™ Polymerase	1 unit
10 pg probe DNA	
10 pmol primer 1	

Note:Use for:

10X dNTP + DIG-11-dUTP (1:5)	< 1 kb
10X dNTP + DIG-11-dUTP (1:10)	1 kb to 1.8 kb
10X dNTP + DIG-11-dUTP (1:15)	> 1.8 kb

2. The PCR reaction uses the following amplification cycles:

- 1) 94°C for 10 min.

2) 5 cycles:	3) 5 cycles:	4) 25 cycles:
95°C - 30 sec 61°C - 1 min 73°C - 5 min	95°C - 30 sec 59°C - 1 min 75°C - 5 min	95°C - 30 sec 51°C - 1 min 73°C - 5 min

- 5) 72°C for 8 min. The reactions are terminated by chilling to 4°C (hold).
3. The products are analyzed by electrophoresis- in a 1% agarose gel, comparing to an aliquot of the unlabelled probe starting material.
4. The amount of DIG-labeled probe is determined as follows:

861

Make serial dilutions of the diluted control DNA in dilution buffer (TE: 10 mM Tris and 1 mM EDTA, pH 8) as shown in the following table:

DIG-labeled control DNA starting conc.	Stepwise Dilution	Final Conc. (Dilution Name)
5 ng/ μ l	1 μ l in 49 μ l TE	100 pg/ μ l (A)
100 pg/ μ l (A)	25 μ l in 25 μ l TE	50 pg/ μ l (B)
50 pg/ μ l (B)	25 μ l in 25 μ l TE	25 pg/ μ l (C)
25 pg/ μ l (C)	20 μ l in 30 μ l TE	10 pg/ μ l (D)

- a. Serial dilutions of a DIG-labeled standard DNA ranging from 100 pg to 10 pg are spotted onto a positively charged nylon membrane, marking the membrane lightly with a pencil to identify each dilution.
- b. Serial dilutions (e.g., 1:50, 1:2500, 1:10,000) of the newly labeled DNA probe are spotted.
- c. The membrane is fixed by UV crosslinking.
- d. The membrane is wetted with a small amount of maleate buffer and then incubated in 1% blocking solution for 15 min at room temp.
- e. The labeled DNA is then detected using alkaline phosphatase conjugated anti-DIG antibody (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) and an NBT substrate according to the manufacture's instruction.
- f. Spot intensities of the control and experimental dilutions are then compared to estimate the concentration of the PCR-DIG-labeled probe.

D. Prehybridization and Hybridization of Southern BlotsSolutions:

100% Formamide purchased from Gibco

20X SSC (1X = 0.15 M NaCl, 0.015 M Na₃citrate)

per L: 175 g NaCl
87.5 g Na₃citrate·2H₂O

20% Sarkosyl (N-lauroyl-sarcosine)

20% SDS (sodium dodecyl sulphate)

10% Blocking Reagent: In 80 ml deionized distilled water, dissolve 1.16 g maleic acid. Next, add NaOH to adjust the pH to 7.5. Add 10 g of the blocking reagent powder. Heat to 60°C while stirring to dissolve the powder. Adjust the volume to 100 ml with water. Stir and sterilize.

Prehybridization Mix:

Final Concentration	Components	Volume (per 100 ml)	Stock
50%	Formamide	50 ml	100%
5X	SSC	25 ml	20X
0.1%	Sarkosyl	0.5 ml	20%
0.02%	SDS	0.1 ml	20%
2%	Blocking Reagent	20 ml	10%
	Water	4.4 ml	

General Procedures:

- Place the blot in a heat-sealable plastic bag and add an appropriate volume of prehybridization solution (30 ml/100cm²) at room temperature. Seal the bag with a heat sealer, avoiding bubbles as much as possible. Lay down the bags in a large plastic tray (one tray can accommodate at least 4–5 bags). Ensure that the bags are

lying flat in the tray so that the prehybridization solution is evenly distributed throughout the bag. Incubate the blot for at least 2 hours with gentle agitation using a waver shaker.

2. Denature DIG-labeled DNA probe by incubating for 10 min. at 98°C using the PCR machine and immediately cool it to 4°C.
3. Add probe to prehybridization solution (25 ng/ml; 30 ml = 750 ng total probe) and mix well but avoid foaming. Bubbles may lead to background.
4. Pour off the prehybridization solution from the hybridization bags and add new prehybridization and probe solution mixture to the bags containing the membrane.
5. Incubate with gentle agitation for at least 16 hours.
6. Proceed to medium stringency post-hybridization wash:

Three times for 20 min. each with gentle agitation using 1X SSC, 1% SDS at 60°C.

All wash solutions must be prewarmed to 60°C. Use about 100 ml of wash solution per membrane.
7. To avoid background keep the membranes fully submerged to avoid drying in spots; agitate sufficiently to avoid having membranes stick to one another.
- After the wash, proceed to immunological detection and CSPD development.

E. Procedure for Immunological Detection with CSPD

Solutions:

- | | | |
|----|-----------|---|
| 20 | Buffer 1: | Maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl; adjusted to pH 7.5 with NaOH) |
|----|-----------|---|

864

Washing buffer: Maleic acid buffer with 0.3% (v/v) Tween 20.

Blocking stock solution 10% blocking reagent in buffer 1. Dissolve (10X concentration): blocking reagent powder (Boehringer Mannheim, Indianapolis, IN, cat. no. 1096176) by constantly stirring on a 65°C heating block or heat in a microwave, autoclave and store at 4°C.

Buffer 2

(1X blocking solution): Dilute the stock solution 1:10 in Buffer 1.

Detection buffer: 0.1 M Tris, 0.1 M NaCl, pH 9.5

Procedure:

1. After the post-hybridization wash the blots are briefly rinsed (1-5 min.) in the maleate washing buffer with gentle shaking.
2. Then the membranes are incubated for 30 min. in Buffer 2 with gentle shaking.
3. Anti-DIG-AP conjugate (Boehringer Mannheim, Indianapolis, IN, cat. no. 1093274) at 75 mU/ml (1:10,000) in Buffer 2 is used for detection. 75 ml of solution can be used for 3 blots.
4. The membrane is incubated for 30 min. in the antibody solution with gentle shaking.
5. The membrane are washed twice in washing buffer with gentle shaking. About 250 mls is used per wash for 3 blots.
6. The blots are equilibrated for 2-5 min in 60 ml detection buffer.
7. Dilute CSPD (1:200) in detection buffer. (This can be prepared ahead of time and stored in the dark at 4°C).

The following steps must be done individually. Bags (one for detection and one for exposure) are generally cut and ready before doing the following steps.

8. The blot is carefully removed from the detection buffer and excess liquid removed without drying the membrane. The blot is immediately placed in a bag and 1.5 ml of CSPD solution is added. The CSPD solution can be spread over the membrane. Bubbles present at the edge and on the surface of the blot are typically removed by gentle rubbing. The membrane is incubated for 5 min. in CSPD solution.
9. Excess liquid is removed and the membrane is blotted briefly (DNA side up) on Whatman 3MM paper. Do not let the membrane dry completely.
10. Seal the damp membrane in a hybridization bag and incubate for 10 min at 37°C to enhance the luminescent reaction.
11. Expose for 2 hours at room temperature to X-ray film. Multiple exposures can be taken. Luminescence continues for at least 24 hours and signal intensity increases during the first hours.

Example 3: Transformation of Carrot Cells

Transformation of plant cells can be accomplished by a number of methods, as described above. Similarly, a number of plant genera can be regenerated from tissue culture following transformation. Transformation and regeneration of carrot cells as described herein is illustrative.

Single cell suspension cultures of carrot (*Daucus carota*) cells are established from hypocotyls of cultivar Early Nantes in B₅ growth medium (O.L. Gamborg et al., *Plant Physiol.* 45:372 (1970)) plus 2,4-D and 15 mM CaCl₂ (B₅-44 medium) by methods known in the art. The suspension cultures are subcultured by adding 10 ml of the suspension culture to 40 ml of B₅-44 medium in 250 ml flasks every 7 days and are maintained in a shaker at 150 rpm at 27 °C in the dark.

The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-

Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM CaCl₂ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution containing 5 mM MES, 20 mM CaCl₂, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4×10^6 protoplasts per ml.

15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Plt. Phys.* 79:988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures*.

An additional deposit, PTA-1411, of an *E. coli* Library, *E. coli*LibA021800, was made at the American Type Culture Collection in Manassas, Virginia, USA on February 22, 2000 to meet the requirements of Budapest Treaty for the international recognition of the deposit of microorganisms. This deposit was assigned ATCC accession no. PTA-1411.

The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

TABLE 1

	>1297184	/22656			
5	len =	1421	nex =	3	
	Term	10090	9717	-	0
	Intr	10506	10184	-	0
	Init	11137	10900	-	0
10	>1297184	/38841			
	len =	1470	nex =	4	
15	Term	14341	13880	-	0
	Intr	14529	14477	-	0
	Intr	14673	14624	-	0
	Init	15349	15056	-	0
20	>1297184	/40037			
	len =	1735	nex =	3	
	Init	16472	16883	+	0
25	Intr	17095	17382	+	0
	Term	17730	18206	+	0
	>1297184	/37635			
30	len =	3070	nex =	11	
	Init	23715	23788	+	0
	Intr	24275	24361	+	0
	Intr	24477	24554	+	0
35	Intr	24641	24834	+	0
	Intr	24949	25090	+	0
	Intr	25275	25355	+	0
	Intr	25618	25746	+	0
	Intr	25852	25929	+	0
40	Intr	26008	26079	+	0
	Intr	26239	26319	+	0
	Term	26416	26618	+	0
	>1402874	/32010			
45	len =	1171	nex =	4	
	Init	65717	66071	+	0
	Intr	66169	66290	+	0
50	Intr	66363	66515	+	0
	Term	66600	66887	+	0
	>1402874	/16813			
55	len =	753	nex =	3	
	Init	78870	78982	+	0
	Intr	79066	79242	+	0
	Term	79311	79622	+	0
60					

	>1402874	/41074			
	len =	850	nex =	3	
5	Init	80005	80250	+	0
	Intr	80334	80486	+	0
	Term	80573	80854	+	0
	>1532162	/42644			
10	len =	1353	nex =	4	
	Init	10117	10395	+	0
	Intr	10519	10718	+	0
15	Intr	10809	11038	+	0
	Term	11112	11469	+	0
	>1532162	/156172			
20	len =	1286	nex =	4	
	Init	10232	10395	+	0
	Intr	10519	10718	+	0
	Intr	10809	11038	+	0
25	Term	11112	11517	+	0
	>1532162	/1415			
30	len =	649	nex =	2	
	Init	11955	12231	+	0
	Term	12334	12603	+	0
	>1532162	/32937			
35	len =	738	nex =	2	
	Init	11982	12231	+	0
40	Term	12334	12719	+	0
	>1532162	/25057			
	len =	2230	nex =	9	
45	Init	21905	22035	+	0
	Intr	22473	22603	+	0
	Intr	22691	22865	+	0
	Intr	22971	23036	+	0
	Intr	23189	23323	+	0
50	Intr	23415	23494	+	0
	Intr	23568	23670	+	0
	Intr	23758	23816	+	0
	Term	23900	24131	+	0
55	>1532162	/20800			
	len =	2174	nex =	9	
	Init	22161	22221	+	0
60	Intr	22473	22603	+	0

869

	Intr	22691	22865	+	0
	Intr	22971	23036	+	0
	Intr	23189	23323	+	0
	Intr	23415	23494	+	0
5	Intr	23568	23670	+	0
	Intr	23758	23816	+	0
	Term	23900	24131	+	0
	>1532162		/33957		
10	len =	2175	nex =	9	
	Init	22161	22221	+	0
	Intr	22473	22603	+	0
15	Intr	22691	22865	+	0
	Intr	22971	23036	+	0
	Intr	23189	23323	+	0
	Intr	23415	23494	+	0
	Intr	23568	23670	+	0
20	Intr	23758	23816	+	0
	Term	23900	24132	+	0
	>1532162		/154048		
25	len =	638	nex =	1	
	Sngl	28309	28946	+	0
	>1532162		/15529		
30	len =	2414	nex =	1	
	Sngl	40592	41742	+	0
35	>1532162		/39051		
	len =	2127	nex =	1	
	Sngl	45263	44433	-	0
40	>1532162		/15968		
	len =	191	nex =	1	
45	Sngl	47765	47575	-	0
	>1532162		/29991		
	len =	2374	nex =	1	
50	Sngl	48703	47562	-	0
	>1532162		/6135		
55	len =	2370	nex =	1	
	Sngl	48703	47566	-	0
60	>1532162		/26043		

870

	len =	1243	nex =	3	
	Term	50762	50243	-	0
5	Intr	51173	51108	-	0
	Init	51485	51286	-	0
	>1532162 /14942				
10	len =	1966	nex =	3	
	Init	52773	52831	+	0
	Intr	52920	53095	+	0
	Term	53208	53601	+	0
15	>1532162 /98909				
	len =	1886	nex =	3	
20	Init	52773	52831	+	0
	Intr	52920	53095	+	0
	Term	53208	53533	+	0
	>1532162 /28026				
25	len =	1675	nex =	4	
	Term	58063	57720	-	0
	Intr	58192	58151	-	0
	Intr	58514	58430	-	0
30	Init	59394	58958	-	0
	>1707006 /26007				
35	len =	359	nex =	2	
	Init	22636	22811	+	0
	Term	22887	22994	+	0
40	>1707006 /26506				
	len =	2124	nex =	6	
	Init	22665	22811	+	0
	Intr	22887	22993	+	0
45	Intr	23509	23571	+	0
	Intr	23932	23963	+	0
	Intr	24280	24374	+	0
	Term	24494	24773	+	0
50	>1707006 /40748				
	len =	3502	nex =	7	
	Init	50053	50271	+	0
55	Intr	50378	50467	+	0
	Intr	51366	51425	+	0
	Intr	51531	51570	+	0
	Intr	51655	51748	+	0
	Intr	51858	51917	+	0
60	Term	52003	52094	+	0

2025 RELEASE UNDER E.O. 14176

5 >1707006 /125567
 len = 1097 nex = 2
 Term 54221 53867 - 0
 Init 54963 54743 - 0
 10 >1707006 /152227
 len = 430 nex = 1
 Sngl 55477 55054 - 0
 15 >1707006 /38063
 len = 1598 nex = 2
 Term 54226 53885 - 0
 Init 55482 54743 - 0
 20 >1707006 /10375
 len = 744 nex = 2
 Init 58320 58692 + 0
 Term 58784 59063 + 0
 25 >1707006 /10617
 len = 815 nex = 3
 Term 62856 62577 - 0
 Intr 63141 62943 - 0
 30 >1707006 /1711
 len = 801 nex = 3
 Term 62856 62591 - 0
 Intr 63141 62943 - 0
 Init 63391 63248 - 0
 35 >1707006 /29818
 len = 760 nex = 3
 Term 62856 62632 - 0
 Intr 63141 62943 - 0
 Init 63391 63248 - 0
 40 >1707006 /40627
 len = 2499 nex = 10
 Init 67998 68163 + 0
 Intr 68353 68506 + 0
 Intr 68592 68875 + 0
 45 >1707006 /40627
 len = 2499 nex = 10
 Init 67998 68163 + 0
 Intr 68353 68506 + 0
 Intr 68592 68875 + 0
 50 >1707006 /40627
 len = 2499 nex = 10
 Init 67998 68163 + 0
 Intr 68353 68506 + 0
 Intr 68592 68875 + 0
 55 >1707006 /40627
 len = 2499 nex = 10
 Init 67998 68163 + 0
 Intr 68353 68506 + 0
 Intr 68592 68875 + 0
 60 >1707006 /40627
 len = 2499 nex = 10
 Init 67998 68163 + 0
 Intr 68353 68506 + 0
 Intr 68592 68875 + 0

872

	Intr	69314	69419	+	0
	Intr	69514	69596	+	0
	Intr	69689	69834	+	0
	Intr	69966	70071	+	0
5	Intr	70216	70275	+	0
	Term	70361	70496	+	0
	>1707006		/101081		
10	len =	976	nex =	1	
	Sngl	75101	74126	-	0
	>1785673		/23693		
15	len =	622	nex =	1	
	Sngl	31275	31896	+	0
20	>1871173		/38610		
	len =	2054	nex =	8	
25	Init	12146	12237	+	0
	Intr	12284	12483	+	0
	Intr	12573	12704	+	0
	Intr	12790	12866	+	0
	Intr	12970	13105	+	0
	Intr	13186	13326	+	0
30	Intr	13424	13482	+	0
	Term	13584	13650	+	0
	>1871173		/10969		
35	len =	1038	nex =	2	
	Term	49483	49183	-	0
	Init	50220	49576	-	0
40	>1871173		/16493		
	len =	3870	nex =	7	
45	Init	50368	50540	+	0
	Intr	50630	50680	+	0
	Intr	51247	51298	+	0
	Intr	51370	51428	+	0
	Intr	51546	51591	+	0
	Intr	52100	52233	+	0
50	Term	54023	54234	+	0
	>1877523		/96448		
55	len =	822	nex =	4	
	Init	105150	105228	+	0
	Intr	105315	105428	+	0
	Intr	105500	105535	+	0
	Term	105630	105953	+	0
60					

>1877523 /2677

len = 670 nex = 1

5 Sngl 21255 20592 - 0

>1877523 /1693

len = 710 nex = 2

10 Term 21139 20646 - 0
 Init 21355 21247 - 0

>1877523 /40042

15 len = 2369 nex = 8

Term 38060 37972 - 0
 Intr 38794 38656 - 0
 20 Intr 38997 38927 - 0
 Intr 39148 39104 - 0
 Intr 39328 39239 - 0
 Intr 39526 39438 - 0
 25 Intr 39689 39614 - 0
 Init 40034 39798 - 0

>1877523 /35733

30 len = 1245 nex = 3

Term 53914 53583 - 0
 Intr 54335 54159 - 0
 Init 54827 54574 - 0

35 >1877523 /34291

len = 809 nex = 1

40 Sngl 61137 61281 + 0

>1877523 /2979

len = 766 nex = 2

45 Init 61146 61281 + 0
 Term 61565 61911 + 0

>1931636 /93598

50 len = 1589 nex = 3

Init 111855 112746 + 0
 Intr 112845 112949 + 0
 55 Term 113156 113443 + 0

>1931636 /40765

len = 1821 nex = 3

60 Term 50015 49475 - 0

874

	Intr	50253	50130	-	0
	Init	51295	50557	-	0
	>1931636	/20637			
5	len =	644	nex =	1	
	Sngl	63596	62953	-	0
10	>1931636	/14648			
	len =	503	nex =	1	
	Sngl	97733	97231	-	0
15	>1946354	/1391			
	len =	4584	nex =	11	
20	Term	12119	11739	-	0
	Intr	12281	12213	-	0
	Intr	12535	12455	-	0
	Intr	12756	12682	-	0
	Intr	13005	12873	-	0
25	Intr	13304	13257	-	0
	Intr	13613	13401	-	0
	Intr	13994	13833	-	0
	Intr	14593	14363	-	0
	Intr	15009	14680	-	0
30	Init	15456	15157	-	0
	>1946354	/7619			
	len =	939	nex =	2	
35	Init	31875	32384	+	0
	Term	32537	32813	+	0
	>1946354	/34999			
40	len =	1078	nex =	2	
	Init	33182	33416	+	0
	Term	33743	34259	+	0
45	>1946354	/39560			
	len =	674	nex =	1	
50	Sngl	41592	42265	+	0
	>1946354	/41046			
	len =	730	nex =	1	
55	Sngl	57609	58323	+	0
	>1946354	/1820			
60	len =	1190	nex =	2	

875

	Term	7729	6909	-	0
	Init	8098	7816	-	0
5	>1946354 /22671				
	len =	1583	nex =	2	
10	Init	83167	83385	+	0
	Term	83523	83614	+	0
	>2062153 /38051				
15	len =	1491	nex =	3	
	Term	15272	14834	-	0
	Intr	15841	15386	-	0
	Init	16324	16026	-	0
20	>2062153 /119458				
	len =	1513	nex =	1	
25	Sngl	16220	16026	-	0
	>2062153 /157474				
	len =	1497	nex =	3	
30	Term	15272	14858	-	0
	Intr	15841	15386	-	0
	Init	16220	16026	-	0
35	>2062153 /30056				
	len =	1520	nex =	3	
	Term	15272	14836	-	0
	Intr	15841	15386	-	0
40	Init	16220	16026	-	0
	>2062153 /42777				
45	len =	1450	nex =	2	
	Term	24390	23947	-	0
	Init	25283	24512	-	0
50	>2062153 /6448				
	len =	1481	nex =	3	
	Term	24390	23947	-	0
	Intr	24955	24512	-	0
55	Init	25427	25053	-	0
	>2062153 /12715				
60	len =	1976	nex =	3	

[illegible]

Term	55961	55118	-	0
Intr	56262	56051	-	0
Init	57093	56580	-	0

5	>2062153	/30003			
	len =	2057	nex =	3	
10	Init	7382	7843	+	0
	Intr	7929	8378	+	0
	Term	8469	8866	+	0
	>2062153	/32293			
15	len =	790	nex =	1	
	Sngl	69530	68750	-	0
20	>2062153	/29750			
	len =	2112	nex =	3	
	Init	76786	77284	+	0
	Intr	77663	77774	+	0
25	Term	77921	78394	+	0
	>2088638	/9398			
30	len =	616	nex =	1	
	Sngl	103573	102958	-	0
	>2088638	/6732			
35	len =	1632	nex =	2	
	Term	17390	16530	-	0
	Init	18161	17822	-	0
40	>2088638	/39048			
	len =	2533	nex =	7	
	Init	24452	24782	+	0
45	Intr	25154	25378	+	0
	Intr	25457	25551	+	0
	Intr	25633	25822	+	0
	Intr	25917	26041	+	0
50	Intr	26186	26401	+	0
	Term	26486	26984	+	0
	>2088638	/33701			
55	len =	1515	nex =	3	
	Term	32027	31685	-	0
	Intr	32312	32109	-	0
	Init	32802	32388	-	0
60	>2088638	/15207			

877

	len =	2110	nex =	4	
	Term	50426	50181	-	0
5	Intr	50656	50514	-	0
	Intr	51540	51487	-	0
	Init	52290	52051	-	0
	>2088638 /5504				
10	len =	2820	nex =	10	
	Term	52859	52504	-	0
	Intr	53066	52943	-	0
15	Intr	53260	53159	-	0
	Intr	53424	53356	-	0
	Intr	53674	53567	-	0
	Intr	53905	53851	-	0
	Intr	54431	54301	-	0
20	Intr	54618	54544	-	0
	Intr	54880	54803	-	0
	Init	55058	54973	-	0
	>2088638 /35056				
25	len =	1510	nex =	1	
	Sngl	70686	69178	-	0
	>2088638 /32440				
30	len =	919	nex =	3	
	Init	80756	80853	+	0
35	Intr	81026	81170	+	0
	Term	81258	81674	+	0
	>2088638 /5046				
40	len =	1647	nex =	3	
	Init	95145	95511	+	0
	Intr	95860	96013	+	0
	Term	96327	96791	+	0
45	>2098816 /31252				
	len =	704	nex =	1	
50	Sngl	35121	35824	+	0
	>2098816 /15292				
	len =	1279	nex =	1	
55	Sngl	39507	40333	+	0
	>2098816 /36730				
60	len =	2135	nex =	9	

878

	Init	43827	44181	+	0
	Intr	44267	44314	+	0
	Intr	44406	44582	+	0
5	Intr	44668	44818	+	0
	Intr	44908	44994	+	0
	Intr	45079	45203	+	0
	Intr	45282	45400	+	0
	Intr	45483	45594	+	0
10	Term	45685	45961	+	0
>2098816		/8716			
15	len =	1090	nex =	5	
	Init	44941	44994	+	0
	Intr	45079	45203	+	0
	Intr	45282	45400	+	0
	Intr	45483	45594	+	0
20	Term	45685	46007	+	0
>2098816		/36216			
25	len =	2338	nex =	6	
	Init	58990	59535	+	0
	Intr	59663	59944	+	0
	Intr	60031	60178	+	0
	Intr	60282	60367	+	0
30	Intr	60894	60971	+	0
	Term	61070	61327	+	0
>2098816		/42713			
35	len =	2280	nex =	6	
	Init	59046	59535	+	0
	Intr	59663	59944	+	0
	Intr	60031	60178	+	0
40	Intr	60282	60367	+	0
	Intr	60894	60971	+	0
	Term	61070	61325	+	0
>2098816		/36286			
45	len =	643	nex =	1	
	Sngl	6052	5410	-	0
>2098816		/38820			
50	len =	756	nex =	1	
	Sngl	6188	5433	-	0
55	>2098816 /38170				
	len =	2445	nex =	6	
60	Term	63428	62916	-	0

879

	Intr	63750	63522	-	0
	Intr	63933	63894	-	0
	Intr	64507	64381	-	0
	Intr	64935	64803	-	0
5	Init	65360	65060	-	0
>2098816 /40254					
10	len =	628	nex =	1	
	Sngl	69008	68744	-	0
>2098816 /17126					
15	len =	811	nex =	2	
	Term	69008	68666	-	0
	Init	69476	69328	-	0
20	>2098816 /122497				
	len =	359	nex =	1	
	Sngl	70110	69752	-	0
25	>2098816 /36543				
	len =	1173	nex =	1	
30	Sngl	77771	76602	-	0
>2098816 /17357					
35	len =	2350	nex =	6	
	Init	88159	88663	+	0
	Intr	88942	89027	+	0
	Intr	89118	89341	+	0
	Intr	89580	89646	+	0
40	Intr	90016	90126	+	0
	Term	90323	90506	+	0
>2098816 /31770					
45	len =	1183	nex =	3	
	Term	90853	90569	-	0
	Intr	91066	90933	-	0
	Init	91751	91546	-	0
50	>2104523 /21952				
	len =	2710	nex =	2	
55	Term	71964	70632	-	0
	Init	73339	72021	-	0
>2104523 /34676					
60	len =	3970	nex =	12	

2025 RELEASE UNDER E.O. 14176

880

	Term	76257	75990	-	0
	Intr	76489	76350	-	0
5	Intr	77518	77169	-	0
	Intr	77737	77610	-	0
	Intr	77956	77828	-	0
	Intr	78109	78031	-	0
	Intr	78360	78197	-	0
10	Intr	78636	78451	-	0
	Intr	78894	78763	-	0
	Intr	79089	78998	-	0
	Intr	79279	79180	-	0
	Init	79954	79652	-	0
15	>2160132	/9002			
	len =	1994	nex =	4	
20	Term	38308	37073	-	0
	Intr	38529	38400	-	0
	Intr	38751	38614	-	0
	Init	39066	38884	-	0
25	>2160132	/18804			
	len =	1584	nex =	1	
	Sngl	60820	59237	-	0
30	>2160132	/21783			
	len =	415	nex =	1	
	Sngl	78298	78712	+	0
35	>2160132	/21416			
	len =	698	nex =	1	
40	Sngl	79001	78304	-	0
	>2160132	/15957			
45	len =	2326	nex =	10	
	Term	88656	88489	-	0
	Intr	88915	88769	-	0
	Intr	89076	89020	-	0
50	Intr	89255	89214	-	0
	Intr	89496	89416	-	0
	Intr	89765	89687	-	0
	Intr	89957	89886	-	0
	Intr	90142	90042	-	0
	Intr	90343	90314	-	0
55	Init	90814	90417	-	0
	>2160132	/41375			
60	len =	2353	nex =	10	

2025 RELEASE UNDER E.O. 14176

881

	Term	88656	88487	-	0
	Intr	88915	88769	-	0
	Intr	89076	89020	-	0
	Intr	89255	89214	-	0
5	Intr	89496	89416	-	0
	Intr	89765	89687	-	0
	Intr	89957	89886	-	0
	Intr	90142	90042	-	0
	Intr	90343	90314	-	0
10	Init	90839	90417	-	0
 >2160155 /17761					
15	len =	3312	nex =	7	
	Term	15073	14667	-	0
	Intr	15288	15145	-	0
	Intr	15853	15737	-	0
	Intr	16067	15930	-	0
20	Intr	16314	16190	-	0
	Intr	17071	16986	-	0
	Init	17685	17547	-	0
 >2160155 /6295					
25	len =	1644	nex =	2	
	Init	44387	45479	+	0
	Term	45550	46030	+	0
30	 >2160155 /22067				
	len =	2477	nex =	7	
35	Term	52597	52279	-	0
	Intr	52839	52684	-	0
	Intr	53020	52925	-	0
	Intr	53302	53110	-	0
	Intr	54097	53469	-	0
40	Intr	54467	54189	-	0
	Init	54755	54540	-	0
 >2160155 /6319					
45	len =	3299	nex =	11	
	Init	60441	60770	+	0
	Intr	61340	61399	+	0
	Intr	61506	61619	+	0
50	Intr	61883	61948	+	0
	Intr	62027	62134	+	0
	Intr	62237	62320	+	0
	Intr	62639	62740	+	0
	Intr	62828	62941	+	0
55	Intr	63016	63096	+	0
	Intr	63191	63310	+	0
	Term	63481	63739	+	0
 >2160155 /17081					
60					

2025 RELEASE UNDER E.O. 14176

882

	len =	1390	nex =	2	
	Term	5724	5111	-	0
5	Init	6498	6293	-	0
	>2160155 /39525				
	len =	2602	nex =	6	
10	Init	7274	7423	+	0
	Intr	7512	7572	+	0
	Intr	7673	7725	+	0
	Intr	7845	7946	+	0
	Intr	8057	8546	+	0
15	Term	8659	9410	+	0
	>2160155 /6642				
	len =	823	nex =	2	
20	Init	76276	76526	+	0
	Term	76743	77098	+	0
	>2160155 /8575				
25	len =	851	nex =	2	
	Init	76277	76526	+	0
	Term	76743	77127	+	0
30	>2160155 /34772				
	len =	1361	nex =	3	
35	Init	7845	7946	+	0
	Intr	8057	8546	+	0
	Term	8659	9205	+	0
	>2160155 /23319				
40	len =	798	nex =	2	
	Term	86139	86067	-	0
	Init	86864	86267	-	0
45	>2182285 /21725				
	len =	2410	nex =	4	
50	Init	10500	10780	+	0
	Intr	11596	11657	+	0
	Intr	12371	12411	+	0
	Term	12536	12907	+	0
55	>2182285 /2118				
	len =	508	nex =	2	
	Init	33841	33970	+	0
60	Term	34088	34262	+	0

884

	Init	22824	22577	-	0
	>2182286 /15161				
5	len =	336	nex =	1	
	Sngl	32955	32620	-	0
	>2182286 /3538				
10	len =	1090	nex =	5	
	Term	56957	56571	-	0
	Intr	57090	57034	-	0
15	Intr	57291	57192	-	0
	Intr	57436	57378	-	0
	Init	57657	57525	-	0
	>2182286 /31705				
20	len =	1400	nex =	1	
	Sngl	61850	60451	-	0
25	>2182287 /13008				
	len =	2683	nex =	13	
	Init	15455	15581	+	0
30	Intr	15687	15748	+	0
	Intr	15834	15911	+	0
	Intr	15991	16066	+	0
	Intr	16164	16234	+	0
	Intr	16347	16448	+	0
35	Intr	16539	16628	+	0
	Intr	16756	16830	+	0
	Intr	17067	17099	+	0
	Intr	17201	17272	+	0
	Intr	17463	17525	+	0
40	Intr	17658	17773	+	0
	Term	17832	18137	+	0
	>2182287 /14016				
45	len =	1166	nex =	1	
	Sngl	33340	33469	+	0
	>2182287 /35042				
50	len =	497	nex =	1	
	Sngl	33706	33927	+	0
	>2182287 /20858				
55	len =	1183	nex =	1	
	Sngl	50135	49403	-	0
60					

885

>2182287 /2985

len = 1215 nex = 2

5	Init	66989	67367	+	0
	Term	67700	68203	+	0

>2182287 /20754

10 len = 1074 nex = 2

	Init	67134	67367	+	0
	Term	67700	68203	+	0

15 >2182287 /13385

len = 3085 nex = 14

	Init	70516	70728	+	0
20	Intr	71073	71225	+	0
	Intr	71312	71410	+	0
	Intr	71569	71670	+	0
	Intr	71750	71893	+	0
	Intr	72009	72103	+	0
25	Intr	72210	72339	+	0
	Intr	72415	72519	+	0
	Intr	72624	72730	+	0
	Intr	72806	72857	+	0
	Intr	72941	73030	+	0
30	Intr	73120	73204	+	0
	Intr	73290	73354	+	0
	Term	73441	73600	+	0

>2182287 /121535

35 len = 153 nex = 1

Sngl 88279 88127 - 0

40 >2182287 /103034

len = 910 nex = 3

	Term	97140	96927	-	0
45	Intr	97606	97230	-	0
	Init	97824	97721	-	0

>2182289 /27197

50 len = 2831 nex = 11

	Term	14016	13844	-	0
	Intr	14182	14117	-	0
	Intr	14330	14282	-	0
55	Intr	14688	14621	-	0
	Intr	14820	14775	-	0
	Intr	15017	14924	-	0
	Intr	15178	15116	-	0
	Intr	15465	15343	-	0
60	Intr	15654	15621	-	0

2025-09-22 14:00:00

886

Intr	16127	16011	-	0
Init	16674	16500	-	0

>2182289 /205500

```

5      len =      1030      nex =          1
      Sngl  52945  51919      -          0

```

10 >2182289 /31372

```
len = 1810    nex = 7
```

	Term	58360	58263	-	0
15	Intr	58757	58601	-	0
	Intr	59020	58889	-	0
	Intr	59232	59120	-	0
	Intr	59487	59328	-	0
	Intr	59636	59613	-	0
20	Init	60069	59921	-	0

>2182289 /38858

```

25      len =      1690      nex =          3
      Term 85015 84307 - 0
      Intr 85533 85226 - 0
      Init 85988 85678 - 0

```

30 >2191126 /28640

```
len = 1810    nex = 3
```

35	Init	105687	105961	+	0
	Intr	106179	106465	+	0
	Term	106570	107490	+	0

>2191126 /1204

```

40      len =      1690      nex =          1
      Snq1 110532 112213      +          0

```

>2191126 /41187

```

len =      2551      nex =      7
Term 115853 115628 - 0
Intr 116345 116037 - 0
Intr 116498 116418 - 0
Intr 116825 116751 - 0
Intr 116990 116904 - 0
Intr 117192 117090 - 0
Init 118178 117485 - 0

```

>2191126 /21195

```

len =      2566    nex =          7
60      Term 115853 115637      -          0

```


887

	Intr	116345	116037	-	0
	Intr	116498	116418	-	0
	Intr	116825	116751	-	0
	Intr	116990	116904	-	0
5	Intr	117192	117090	-	0
	Init	118202	117485	-	0

>2191126 /19141

10	len =	3438	nex =	14	
	Term	25095	24742	-	0
	Intr	25245	25180	-	0
	Intr	25409	25338	-	0
15	Intr	25625	25512	-	0
	Intr	25812	25720	-	0
	Intr	25961	25899	-	0
	Intr	26152	26042	-	0
	Intr	26360	26247	-	0
20	Intr	26604	26506	-	0
	Intr	26756	26691	-	0
	Intr	26948	26853	-	0
	Intr	27119	27035	-	0
	Intr	27350	27203	-	0
25	Init	28179	28046	-	0

>2191126 /22919

	len =	1497	nex =	4	
30	Init	28448	28746	+	0
	Intr	29035	29235	+	0
	Intr	29321	29463	+	0
	Term	29617	29944	+	0

>2191126 /117191

	len =	253	nex =	1	
40	Sngl	66070	66322	+	0

>2191126 /7653

	len =	1991	nex =	6	
45	Term	5264	4896	-	0
	Intr	5520	5337	-	0
	Intr	5933	5601	-	0
	Intr	6260	6123	-	0
50	Intr	6458	6329	-	0
	Init	6886	6590	-	0

>2191126 /41270

55	len =	1008	nex =	4	
	Term	79378	79212	-	0
	Intr	79532	79461	-	0
	Intr	79730	79623	-	0
60	Init	79919	79819	-	0

00190.000000

888

>2191126 /94836

5	len =	1131	nex =	5	
	Term	79378	79190	-	0
	Intr	79532	79461	-	0
	Intr	79730	79623	-	0
	Intr	79919	79819	-	0
10	Init	80320	80142	-	0

>2191126 /12604

15	len =	1177	nex =	5	
	Term	79378	79212	-	0
	Intr	79532	79461	-	0
	Intr	79730	79623	-	0
	Intr	79919	79819	-	0
20	Init	80388	80142	-	0

>2191126 /16533

25	len =	2830	nex =	12	
	Init	90009	90076	+	0
	Intr	90175	90222	+	0
	Intr	90315	90410	+	0
	Intr	90483	90563	+	0
30	Intr	90659	90700	+	0
	Intr	90784	90947	+	0
	Intr	91030	91078	+	0
	Intr	91166	91214	+	0
	Intr	91306	91445	+	0
35	Intr	91538	91615	+	0
	Intr	91709	91834	+	0
	Term	91909	92323	+	0

>2191157 /5457

40	len =	688	nex =	2	
	Term	110545	110202	-	0
45	Init	110889	110723	-	0

>2191157 /39714

	len =	520	nex =	1	
50	Sngl	24526	25045	+	0

>2191157 /37336

55	len =	1558	nex =	2	
	Init	26629	26769	+	0
	Term	27064	27170	+	0

>2191157 /17739

60

2025-09-25 09:09:09

889

	len =	2326	nex =	11	
	Term	1098	904	-	0
5	Intr	1303	1201	-	0
	Intr	1501	1418	-	0
	Intr	1698	1603	-	0
	Intr	1848	1798	-	0
	Intr	2076	1936	-	0
	Intr	2220	2164	-	0
10	Intr	2391	2317	-	0
	Intr	2739	2467	-	0
	Intr	2894	2835	-	0
	Init	3094	3002	-	0
15	>2191157		/21258		
	len =	2364	nex =	9	
20	Init	35554	35767	+	0
	Intr	35854	35917	+	0
	Intr	36017	36231	+	0
	Intr	36362	36538	+	0
	Intr	36622	36696	+	0
25	Intr	36794	36895	+	0
	Intr	37265	37376	+	0
	Intr	37474	37620	+	0
	Term	37753	37793	+	0
	>2191157		/42174		
30	len =	540	nex =	1	
	Sngl	59287	59826	+	0
35	>2191157		/27625		
	len =	732	nex =	2	
40	Init	80900	81166	+	0
	Term	81274	81631	+	0
	>2191157		/41361		
45	len =	2136	nex =	2	
	Init	83526	83731	+	0
	Term	83861	84187	+	0
	>2191157		/32265		
50	len =	2008	nex =	2	
	Init	83526	83731	+	0
	Term	83861	84181	+	0
55	>2191157		/2495		
	len =	2795	nex =	5	
60	Init	92543	92875	+	0

2025 RELEASE UNDER E.O. 14176

890

	Intr	93634	93776	+	0
	Intr	94054	94077	+	0
	Intr	94512	94714	+	0
	Term	94965	95337	+	0
5	>2191181 /38304				
	len =	2070	nex =	4	
10	Init	1742	2050	+	0
	Intr	2468	2686	+	0
	Intr	2758	2844	+	0
	Term	3193	3219	+	0
15	>2191181 /23239				
	len =	988	nex =	3	
	Term	4337	3802	-	0
20	Intr	4497	4418	-	0
	Init	4789	4601	-	0
	>2191181 /30935				
25	len =	1455	nex =	0	
	>2213606 /6503				
	len =	1974	nex =	4	
30	Init	15815	16171	+	0
	Intr	16373	16842	+	0
	Intr	16925	17188	+	0
	Term	17281	17788	+	0
35	>2213606 /10990				
	len =	413	nex =	1	
40	Sngl	18252	17840	-	0
	>2213606 /38093				
	len =	490	nex =	1	
45	Sngl	27032	27514	+	0
	>2213606 /23231				
50	len =	700	nex =	1	
	Sngl	45292	44593	-	0
	>2213606 /31944				
55	len =	559	nex =	1	
	Sngl	49930	49372	-	0
60	>2244747 /16846				

2025 RELEASE UNDER E.O. 14176

	len =	2017	nex =	2	
	Init	12786	13565	+	0
5	Term	13854	14802	+	0
	>2244747 /38987				
	len =	134	nex =	1	
10	Sngl	14762	14895	+	0
	>2244747 /17977				
15	len =	610	nex =	1	
	Sngl	16599	15997	-	0
	>2244747 /19172				
20	len =	610	nex =	1	
	Sngl	16614	16009	-	0
25	>2244747 /30129				
	len =	813	nex =	1	
	Sngl	176792	177114	+	0
30	>2244747 /195				
	len =	805	nex =	1	
35	Sngl	176309	177113	+	0
	>2244747 /101734				
40	len =	340	nex =	1	
	Sngl	198899	199238	+	0
	>2244747 /126389				
45	len =	1776	nex =	8	
	Init	48741	48903	+	0
	Intr	48995	49057	+	0
	Intr	49141	49207	+	0
50	Intr	49296	49396	+	0
	Intr	49486	49530	+	0
	Intr	49614	49895	+	0
	Intr	49983	50085	+	0
	Term	50189	50516	+	0
55	>2244747 /25991				
	len =	1850	nex =	8	
60	Init	48741	48903	+	0

892

	Intr	48995	49057	+	0
	Intr	49141	49207	+	0
	Intr	49296	49396	+	0
	Intr	49486	49530	+	0
5	Intr	49614	49895	+	0
	Intr	49983	50085	+	0
	Term	50189	50590	+	0
	>2244747 /99093				
10	len =	430	nex =	3	
	Init	48743	48903	+	0
	Intr	48995	49057	+	0
15	Term	49141	49172	+	0
	>2244747 /7346				
	len =	550	nex =	1	
20	Sngl	51305	50761	-	0
	>2244747 /13520				
25	len =	522	nex =	1	
	Sngl	53660	53139	-	0
	>2244747 /18697				
30	len =	817	nex =	2	
	Term	56326	55871	-	0
	Init	56687	56413	-	0
35	>2244747 /35186				
	len =	1525	nex =	5	
40	Term	56326	55870	-	0
	Intr	56685	56413	-	0
	Intr	56884	56777	-	0
	Intr	57220	56989	-	0
	Init	57394	57303	-	0
45	>2244747 /39975				
	len =	2277	nex =	7	
50	Term	56326	55859	-	0
	Intr	56685	56413	-	0
	Intr	56884	56777	-	0
	Intr	57220	56989	-	0
	Intr	57530	57303	-	0
55	Intr	57816	57621	-	0
	Init	58135	57936	-	0
	>2244747 /108308				
60	len =	2306	nex =	6	

893

	Term	58896	58494	-	0
	Intr	59256	58984	-	0
	Intr	59446	59412	-	0
5	Intr	59994	59535	-	0
	Term	60270	60075	-	0
	Init	60799	60608	-	0
	>2244747 /34967				
10	len =	1692	nex =	3	
	Init	78644	78978	+	0
	Intr	79811	79967	+	0
15	Term	80055	80335	+	0
	>2244747 /29662				
	len =	2324	nex =	4	
20	Term	6181	5707	-	0
	Intr	6376	6275	-	0
	Intr	6858	6468	-	0
	Init	8030	7268	-	0
25	>2244747 /10852				
	len =	948	nex =	3	
30	Term	95484	95087	-	0
	Intr	95756	95563	-	0
	Init	96034	95845	-	0
	>2244747 /33554				
35	len =	1225	nex =	3	
	Term	95484	94981	-	0
	Intr	95756	95563	-	0
40	Init	96205	95845	-	0
	>2244788 /33860				
	len =	894	nex =	2	
45	Init	119066	119340	+	0
	Term	119433	119959	+	0
	>2244788 /4232				
50	len =	1570	nex =	3	
	Term	11837	11610	-	0
	Intr	12997	12874	-	0
55	Init	13171	13086	-	0
	>2244788 /20129				
60	len =	1736	nex =	4	

894

	Init	134496	134633	+	0
	Intr	134785	134908	+	0
	Intr	135250	135306	+	0
	Term	135918	136231	+	0
5	>2244788 /4905				
	len =	1532	nex =	4	
10	Init	134547	134633	+	0
	Intr	134785	134908	+	0
	Intr	135250	135306	+	0
	Term	135918	136078	+	0
15	>2244788 /18255				
	len =	1917	nex =	4	
	Term	11837	11553	-	0
20	Intr	12997	12874	-	0
	Intr	13171	13086	-	0
	Init	13469	13401	-	0
	>2244788 /42223				
25	len =	1270	nex =	2	
	Init	141770	141970	+	0
	Term	142713	143034	+	0
30	>2244788 /21908				
	len =	865	nex =	2	
35	Term	172609	172540	-	0
	Init	173404	172806	-	0
	>2244788 /95834				
40	len =	932	nex =	4	
	Init	176283	176507	+	0
	Intr	176602	176703	+	0
	Intr	176785	176939	+	0
45	Term	176951	177214	+	0
	>2244788 /31495				
	len =	1150	nex =	5	
50	Init	177820	177887	+	0
	Intr	178110	178208	+	0
	Intr	178295	178347	+	0
	Intr	178445	178518	+	0
55	Term	178797	178969	+	0
	>2244788 /40073				
60	len =	1761	nex =	5	

895

	Term	182960	182681	-	0
	Intr	183144	183074	-	0
	Intr	183352	183228	-	0
	Intr	183544	183430	-	0
5	Init	184441	183731	-	0

>2244788 /2738

```

10      len =      1855      nex =      7
      Term 182960 182701      -      0
      Intr 183144 183074      -      0
      Intr 183352 183228      -      0
      Intr 183544 183430      -      0
15      Intr 183825 183731      -      0
      Intr 184012 183901      -      0
      Init 184555 184343      -      0

```

>2244788 /18153

```

len =      1337      nex =          4
Term      744        549      -        0
Intr      903        829      -        0
Intr     1232       1053      -        0
Init     1885       1804      -        0

```

>2244788 /16319

```

30      len =      1732      nex =      7
      Term 188526 188214      -      0
      Intr 188710 188640      -      0
35      Intr 188914 188790      -      0
      Intr 189112 188998      -      0
      Intr 189340 189246      -      0
      Intr 189532 189421      -      0
      Init 189945 189850      -      0

```

40 >2244788 /34477

	len =	790	nex =	4	
45	Term	26188	26035	-	0
	Intr	26496	26276	-	0
	Intr	26702	26590	-	0
	Init	26822	26779	-	0

>2244788 /37809

	len =	2215	nex =	10	
	Term	29960	29503	-	0
55	Intr	30139	30054	-	0
	Intr	30309	30235	-	0
	Intr	30490	30388	-	0
	Intr	30687	30606	-	0
	Intr	30881	30790	-	0
	Intr	31057	30969	-	0
60	Intr	31236	31156	-	0

896

Intr	31450	31336	-	0
Init	31717	31579	-	0

>2244788 /9870

5	len =	1700	nex =	6	
	Term	45280	45046	-	0
	Intr	45431	45380	-	0
10	Intr	45545	45518	-	0
	Intr	46149	46080	-	0
	Intr	46413	46313	-	0
	Init	46745	46519	-	0

15 >2244788 /40736

	len =	1713	nex =	5	
	Init	57948	58133	+	0
20	Intr	58560	58765	+	0
	Intr	58850	58930	+	0
	Intr	59012	59174	+	0
	Term	59262	59660	+	0

25 >2244788 /1718

	len =	1844	nex =	5	
	Term	60276	59985	-	0
30	Intr	60467	60369	-	0
	Intr	60644	60555	-	0
	Intr	60856	60742	-	0
	Init	61828	61672	-	0

35 >2244788 /94503

	len =	1930	nex =	5	
	Term	60276	59949	-	0
40	Intr	60467	60369	-	0
	Intr	60644	60555	-	0
	Intr	60856	60742	-	0
	Init	61875	61672	-	0

45 >2244788 /28978

	len =	921	nex =	1	
	Sngl	63706	62786	-	0

50 >2244788 /36844

	len =	1309	nex =	1	
55	Sngl	78815	80123	+	0

>2244788 /42933

60	len =	2960	nex =	6	
----	-------	------	-------	---	--

897

	Init	92232	92765	+	0
	Intr	92959	93121	+	0
	Intr	93567	93743	+	0
	Intr	93831	93914	+	0
5	Intr	94438	94519	+	0
	Term	94602	95191	+	0

>2244829 /38042

10	len =	2717	nex =	9	
	Init	103735	104049	+	0
	Intr	104329	104423	+	0
	Intr	104545	104609	+	0
15	Intr	104833	104876	+	0
	Intr	105212	105295	+	0
	Intr	105486	105639	+	0
	Intr	105738	105920	+	0
20	Intr	106013	106069	+	0
	Term	106159	106451	+	0

>2244829 /293

25	len =	315	nex =	1	
	Sngl	114012	113698	-	0

>2244829 /40074

30	len =	1498	nex =	2	
	Term	115095	113973	-	0
	Init	115470	115294	-	0

35	>2244829	/38411			
	len =	1796	nex =	2	
	Term	115095	113698	-	0
40	Init	115493	115294	-	0

>2244829 /10518

45	len =	2190	nex =	8	
	Init	116378	116531	+	0
	Intr	116787	116872	+	0
	Intr	116953	117024	+	0
	Intr	117143	117180	+	0
50	Intr	117526	117569	+	0
	Intr	117791	117837	+	0
	Intr	117992	118166	+	0
	Term	118269	118567	+	0

55	>2244829	/29288			
	len =	492	nex =	1	
60	Sngl	131227	130736	-	0

2025 RELEASE UNDER E.O. 14176

	>2244829	/24175		
	len =	332	nex =	1
5	Sngl	136899	137230	+ 0
	>2244829	/17179		
	len =	450	nex =	1
10	Sngl	136899	137332	+ 0
	>2244829	/99523		
	len =	346	nex =	1
15	Sngl	136900	137245	+ 0
	>2244829	/37184		
20	len =	624	nex =	0
	>2244829	/126602		
25	len =	654	nex =	1
	Sngl	136900	137553	+ 0
	>2244829	/15384		
30	len =	627	nex =	1
	Sngl	136904	137530	+ 0
35	>2244829	/26797		
	len =	628	nex =	1
40	Sngl	136904	137531	+ 0
	>2244829	/36129		
	len =	739	nex =	1
45	Sngl	199828	200566	+ 0
	>2244829	/24266		
	len =	1908	nex =	3
50	Init	65354	65621	+ 0
	Intr	65713	65836	+ 0
	Term	66807	67261	+ 0
55	>2244829	/31856		
	len =	897	nex =	3
	Init	70117	70500	+ 0
60	Intr	70585	70611	+ 0

899

	Term	70696	71013	+	0
	>2244829 /30327				
5	len =	711	nex =	1	
	Sngl	82258	82968	+	0
	>2244829 /33166				
10	len =	650	nex =	1	
	Sngl	82303	82952	+	0
	>2244829 /42848				
15	len =	2473	nex =	9	
	Term	83367	83062	-	0
20	Intr	83556	83476	-	0
	Intr	83703	83644	-	0
	Intr	83890	83811	-	0
	Intr	84071	84020	-	0
	Intr	84306	84169	-	0
25	Intr	84661	84398	-	0
	Intr	84799	84742	-	0
	Init	84996	84887	-	0
	>2244829 /22861				
30	len =	611	nex =	1	
	Sngl	85902	86512	+	0
	>2244829 /25333				
35	len =	2115	nex =	3	
	Term	87340	86629	-	0
40	Intr	87618	87443	-	0
	Init	88743	87767	-	0
	>2244829 /117350				
45	len =	1760	nex =	8	
	Term	93545	93422	-	0
	Intr	93819	93710	-	0
	Intr	93998	93936	-	0
50	Intr	94168	94094	-	0
	Intr	94368	94276	-	0
	Intr	94573	94469	-	0
	Intr	94861	94740	-	0
	Init	95181	94950	-	0
55	>2244870 /2163				
	len =	1517	nex =	1	
60	Sngl	13507	15023	+	0

ORIGINAL DOCUMENT

900

>2244870 /15641

5	len =	1853	nex =	7		
	Init	2352	2569	+	0	
	Intr	2668	2781	+	0	
	Intr	2862	2957	+	0	
	Intr	3057	3099	+	0	
10	Intr	3174	3326	+	0	
	Intr	3408	3476	+	0	
	Term	3843	4204	+	0	

>2244870 /35290

15	len =	1090	nex =	2		
	Term	33366	33045	-	0	
	Init	34113	33943	-	0	

>2244870 /18642

20	len =	867	nex =	2		
25	Term	4431	4071	-	0	
	Init	4937	4513	-	0	

>2244870 /30852

30	len =	513	nex =	1		
	Sngl	70945	70433	-	0	

>2244870 /36205

35	len =	1210	nex =	1		
	Sngl	71644	70435	-	0	

40 >2244870 /30929

	len =	867	nex =	1		
	Sngl	84563	85414	+	0	

45 >2244901 /32219

	len =	644	nex =	1		
--	-------	-----	-------	---	--	--

50 Sngl 100297 100940 + 0

>2244901 /101301

	len =	1235	nex =	2		
--	-------	------	-------	---	--	--

55	Init	12251	12597	+	0	
	Term	13371	13485	+	0	

>2244901 /15334

60

00505726-051601

901

	len =	2089	nex =	4	
	Init	12251	12597	+	0
	Intr	13371	13484	+	0
5	Intr	13678	13835	+	0
	Term	13944	14339	+	0
	>2244901 /14485				
10	len =	1048	nex =	2	
	Term	136645	136202	-	0
	Init	137249	136976	-	0
15	>2244901 /8916				
	len =	761	nex =	2	
	Init	146636	146871	+	0
20	Term	146912	147396	+	0
	>2244901 /22637				
	len =	1930	nex =	7	
25	Init	150934	151112	+	0
	Intr	151807	151845	+	0
	Intr	151938	151991	+	0
	Intr	152091	152144	+	0
30	Intr	152269	152322	+	0
	Intr	152417	152488	+	0
	Term	152622	152862	+	0
	>2244901 /5455				
35	len =	550	nex =	1	
	Sngl	153514	154059	+	0
40	>2244901 /25390				
	len =	1731	nex =	3	
	Term	156239	156216	-	0
45	Intr	156385	156332	-	0
	Init	157099	156997	-	0
	>2244901 /39757				
50	len =	1489	nex =	5	
	Term	164193	163773	-	0
	Intr	164487	164293	-	0
	Intr	164750	164603	-	0
55	Intr	164938	164832	-	0
	Init	165261	165017	-	0
	>2244901 /113295				
60	len =	250	nex =	1	

2025 RELEASE UNDER E.O. 14176

902

	Sngl	165261	165021	-	0
	>2244901	/43007			
5	len =	3418	nex =	9	
	Init	181307	182180	+	0
	Intr	182482	182558	+	0
10	Intr	182639	182732	+	0
	Intr	182817	182915	+	0
	Intr	183212	183301	+	0
	Intr	183400	183519	+	0
	Intr	183767	183870	+	0
15	Intr	184163	184235	+	0
	Term	184397	184724	+	0
	>2244901	/8381			
20	len =	928	nex =	2	
	Init	197128	197392	+	0
	Term	197699	198055	+	0
25	>2244901	/35383			
	len =	1690	nex =	1	
	Sngl	23032	21343	-	0
30	>2244901	/12451			
	len =	2050	nex =	4	
35	Init	29261	29459	+	0
	Intr	29681	29785	+	0
	Intr	29969	30397	+	0
	Term	30959	31303	+	0
40	>2244901	/8234			
	len =	855	nex =	4	
	Term	33518	33296	-	0
45	Intr	33802	33633	-	0
	Intr	34017	33880	-	0
	Init	34150	34103	-	0
	>2244901	/33073			
50	len =	3028	nex =	2	
	Init	4164	4631	+	0
	Term	6071	7191	+	0
55	>2244901	/307			
	len =	1838	nex =	8	
60	Init	44565	44888	+	0

903

	Intr	44976	45044	+	0
	Intr	45145	45198	+	0
	Intr	45288	45327	+	0
	Intr	45414	45512	+	0
5	Intr	45595	45819	+	0
	Intr	45902	46023	+	0
	Term	46120	46402	+	0
	>2244901		/19122		
10	len =	1766	nex =	8	
	Init	44638	44888	+	0
	Intr	44976	45044	+	0
15	Intr	45145	45198	+	0
	Intr	45288	45327	+	0
	Intr	45414	45512	+	0
	Intr	45595	45819	+	0
	Intr	45902	46023	+	0
20	Term	46120	46403	+	0
	>2244901		/37345		
25	len =	1379	nex =	3	
	Init	55027	55308	+	0
	Intr	55387	55671	+	0
	Term	55759	56179	+	0
30	>2244901		/26019		
	len =	1750	nex =	3	
	Init	77747	78039	+	0
35	Intr	78780	78906	+	0
	Term	79065	79492	+	0
	>2244901		/933		
40	len =	1415	nex =	2	
	Init	86075	86413	+	0
	Term	86998	87489	+	0
45	>2244950		/12629		
	len =	3346	nex =	10	
	Term	100982	100625	-	0
50	Intr	101466	101106	-	0
	Intr	101718	101591	-	0
	Intr	102002	101874	-	0
	Intr	102439	102360	-	0
	Intr	102690	102527	-	0
55	Intr	102958	102773	-	0
	Intr	103205	103074	-	0
	Intr	103432	103291	-	0
	Init	103970	103568	-	0
60	>2244950		/40414		

2025 RELEASE UNDER E.O. 14176

	len =	2150	nex =	6	
5	Term	109338	109067	-	0
	Intr	109551	109489	-	0
	Intr	109708	109646	-	0
	Intr	109850	109803	-	0
	Intr	110001	109939	-	0
	Init	111043	110961	-	0
10	>2244950 /30227				
	len =	2050	nex =	6	
15	Term	109338	109187	-	0
	Intr	109551	109489	-	0
	Intr	109708	109646	-	0
	Intr	109850	109803	-	0
	Intr	110001	109939	-	0
20	Init	111043	110961	-	0
	>2244950 /5714				
	len =	1403	nex =	7	
25	Init	124186	124326	+	0
	Intr	124418	124469	+	0
	Intr	124596	124670	+	0
	Intr	124766	124794	+	0
30	Intr	124968	125001	+	0
	Intr	125082	125152	+	0
	Term	125251	125588	+	0
	>2244950 /33513				
35	len =	1593	nex =	4	
	Init	138127	138644	+	0
	Intr	138739	138858	+	0
40	Intr	138934	139180	+	0
	Term	139256	139719	+	0
	>2244950 /19028				
45	len =	638	nex =	2	
	Init	139024	139180	+	0
	Term	139256	139661	+	0
50	>2244950 /21894				
	len =	1030	nex =	1	
55	Sngl	146832	145803	-	0
	>2244950 /7605				
	len =	814	nex =	2	
60	Term	167332	166714	-	0

905

Tnit	167527	167451	-	0
------	--------	--------	---	---

>2244950 /3176

```
5      len =      1423      nex =          2
```

Term	167332	166764	-	0
------	--------	--------	---	---

Init	167934	167451	-	0
------	--------	--------	---	---

10 >2244950 /41791

```
len = 1479    nex = 3
```

Term	167332	166712	-	0
------	--------	--------	---	---

15	Intr	167934	167451	-	0
----	------	--------	--------	---	---

Init	168190	168116	-	0
------	--------	--------	---	---

>2244950 /12256

```
20      len =      1716      nex =          4
```

Term	169269	169015	-	0
------	--------	--------	---	---

Intr	169606	169448	-	0
------	--------	--------	---	---

Intr	170335	170260	-	0
------	--------	--------	---	---

Tnit	170730	170607	-	0
------	--------	--------	---	---

>2244950 /6723

```
len = 1536    nex = 4
```

30

Init	171676	171958	+	0
------	--------	--------	---	---

Intr	172224	172415	+	0
------	--------	--------	---	---

Intr	172496	172661	+	0
------	--------	--------	---	---

>2244950 /124835

```
len = 978    nex = 1
```

40 Snq1 18831 19808 + 0

>2244950 /40793

```
len = 1247    nex = 3
```

Term	193189	192906	-	0
------	--------	--------	---	---

Intr	193587	193266	-	0
------	--------	--------	---	---

Tnit	194152	193673	-	0
------	--------	--------	---	---

50 >2244950 /2803

```
len = 1824    nex = 3
```

Init	2896	3184	+	0
------	------	------	---	---

55	Intr	3571	3676	+	0
----	------	------	------	---	---

Term	4403	4719	+	0
------	------	------	---	---

>2244950 / 9209

```
60      len =      573      nex =      1
```

906

	Sngl	31137	30565	-	0
	>2244950	/29655			
5	len =	682	nex =	1	
	Sngl	34486	35167	+	0
10	>2244950	/40913			
	len =	2079	nex =	7	
	Init	4949	5128	+	0
15	Intr	5254	5419	+	0
	Intr	5498	5550	+	0
	Intr	5911	5973	+	0
	Intr	6366	6416	+	0
	Intr	6516	6630	+	0
20	Term	6687	7027	+	0
	>2244950	/18234			
	len =	1950	nex =	6	
25	Init	61059	61335	+	0
	Intr	61420	61550	+	0
	Intr	61714	61791	+	0
	Intr	61882	61926	+	0
30	Intr	62016	62060	+	0
	Term	62293	62389	+	0
	>2244950	/32203			
35	len =	1510	nex =	6	
	Init	7376	7454	+	0
	Intr	7542	7577	+	0
	Intr	7707	7844	+	0
40	Intr	7939	8012	+	0
	Intr	8418	8486	+	0
	Term	8556	8884	+	0
	>2244950	/31782			
45	len =	1211	nex =	1	
	Sngl	84183	82973	-	0
50	>2244950	/17019			
	len =	2897	nex =	2	
	Term	84672	82981	-	0
55	Init	85877	85235	-	0
	>2244950	/109560			
60	len =	397	nex =	1	

2025 RELEASE UNDER E.O. 14176

907

	Sngl	95604	96000	+	0
	>2244991 /7101				
5	len =	1300	nex =	5	
	Term	99473	99160	-	0
	Intr	99674	99597	-	0
	Intr	99851	99788	-	0
10	Intr	100015	99939	-	0
	Init	100216	100170	-	0
	>2244991 /14136				
15	len =	1251	nex =	1	
	Sngl	133001	131751	-	0
	>2244991 /24611				
20	len =	1275	nex =	6	
	Init	144816	144916	+	0
	Intr	144996	145065	+	0
25	Intr	145153	145209	+	0
	Intr	145299	145360	+	0
	Intr	145408	145507	+	0
	Term	145593	145964	+	0
30	>2244991 /5546				
	len =	1163	nex =	3	
	Term	157187	156808	-	0
35	Intr	157430	157305	-	0
	Init	157970	157545	-	0
	>2244991 /8212				
40	len =	1254	nex =	0	
	>2244991 /40778				
	len =	879	nex =	3	
45	Init	163368	163492	+	0
	Intr	163658	163757	+	0
	Term	163863	164240	+	0
50	>2244991 /23771				
	len =	1377	nex =	4	
	Term	164902	164507	-	0
55	Intr	165186	164989	-	0
	Intr	165666	165500	-	0
	Init	165883	165813	-	0
	>2244991 /16525				
60					

908

	len =	810	nex =	2	
	Init	172277	172503	+	0
5	Term	172604	173086	+	0
	>2244991	/22084			
	len =	1450	nex =	2	
10	Init	177203	177333	+	0
	Term	177407	177827	+	0
	>2244991	/157870			
15	len =	342	nex =	1	
	Sngl	17882	17541	-	0
	>2244991	/5686			
20	len =	2453	nex =	10	
	Term	194540	194396	-	0
	Intr	194759	194680	-	0
25	Intr	194888	194843	-	0
	Intr	195027	194971	-	0
	Intr	195163	195105	-	0
	Intr	195344	195244	-	0
	Intr	195623	195502	-	0
30	Intr	195980	195929	-	0
	Intr	196138	196058	-	0
	Init	196848	196213	-	0
	>2244991	/2505			
35	len =	623	nex =	1	
	Sngl	27093	26471	-	0
40	>2244991	/7632			
	len =	1210	nex =	3	
	Term	36794	36385	-	0
45	Intr	37205	37073	-	0
	Init	37590	37308	-	0
	>2244991	/30471			
50	len =	1883	nex =	6	
	Term	39363	38946	-	0
	Intr	39486	39437	-	0
	Intr	39651	39570	-	0
55	Intr	39806	39736	-	0
	Intr	40168	40098	-	0
	Init	40371	40292	-	0
	>2244991	/17535			
60					

	len =	585	nex =	1	
	Sngl	43288	43872	+	0
5	>2244991	/17553			
	len =	628	nex =	2	
10	Init	44575	44786	+	0
	Term	44876	45202	+	0
	>2244991	/16090			
15	len =	634	nex =	2	
	Init	44583	44786	+	0
	Term	44876	45216	+	0
	>2244991	/31946			
20	len =	562	nex =	1	
	Sngl	66524	65963	-	0
25	>2244991	/6580			
	len =	509	nex =	1	
	Sngl	70265	69757	-	0
30	>2244991	/17851			
	len =	1752	nex =	5	
35	Term	71484	71210	-	0
	Intr	71754	71636	-	0
	Intr	71898	71846	-	0
	Intr	72484	72429	-	0
	Init	72626	72579	-	0
40	>2244991	/92054			
	len =	587	nex =	1	
45	Sngl	8564	9150	+	0
	>2245031	/92144			
	len =	444	nex =	1	
50	Sngl	125198	125641	+	0
	>2245031	/30087			
55	len =	822	nex =	1	
	Sngl	125198	126019	+	0
60	>2245031	/118011			

910

```
len =      355    nex =      1
Snq1 125287 125641      +      0
```

5 >2245031 /91870

```
len =      1970      nex =      4

Init 144106 144256      +      0
Intr 144641 144768      +      0
Intr 145143 145253      +      0
Term 145583 146075      +      0
```

>2245031 /36017

```

15      len =      3647      nex =      8
      Term 154141 153926 - 0
      Intr 155021 154948 - 0
20      Intr 155252 155139 - 0
      Intr 155661 155584 - 0
      Intr 155955 155829 - 0
      Intr 156204 156149 - 0
      Intr 156561 156358 - 0
25      Init 157572 157241 - 0

```

>2245031 /7834

	len =	3010	nex =	12	
30	Init	157780	157908	+	0
	Intr	157993	158125	+	0
	Intr	158517	158604	+	0
	Intr	158708	158784	+	0
35	Intr	159068	159107	+	0
	Intr	159412	159497	+	0
	Intr	159590	159671	+	0
	Intr	159798	159854	+	0
40	Intr	159938	159976	+	0
	Intr	160067	160137	+	0
	Intr	160354	160407	+	0
	Term	160554	160780	+	0

>2245031 /114540

45	len =	3018	nex =	11	
	Init	157780	157908	+	0
	Intr	157993	158125	+	0
50	Intr	158517	158604	+	0
	Intr	158708	158784	+	0
	Intr	159068	159497	+	0
	Intr	159590	159671	+	0
	Intr	159798	159854	+	0
55	Intr	159938	159976	+	0
	Intr	160067	160137	+	0
	Intr	160354	160407	+	0
	Term	160554	160797	+	0

60 >2245031 /110681

	len =	466	nex =	2	
	Init	172709	172801	+	0
5	Term	172906	173174	+	0
	>2245031 /142850				
	len =	610	nex =	1	
10	Sngl	173847	173242	-	0
	>2245031 /42533				
15	len =	1533	nex =	4	
	Init	17415	17660	+	0
	Intr	17764	18062	+	0
	Intr	18331	18410	+	0
20	Term	18499	18947	+	0
	>2245031 /36882				
	len =	2299	nex =	5	
25	Term	173963	173241	-	0
	Intr	174262	174007	-	0
	Intr	174516	174406	-	0
	Intr	174824	174614	-	0
30	Init	175539	174923	-	0
	>2245031 /14613				
	len =	673	nex =	1	
35	Sngl	20501	19829	-	0
	>2245031 /831				
40	len =	850	nex =	3	
	Init	39954	40111	+	0
	Intr	40198	40248	+	0
	Term	40330	40796	+	0
45	>2245031 /14223				
	len =	638	nex =	1	
50	Sngl	43095	43370	+	0
	>2245031 /35772				
	len =	1663	nex =	1	
55	Sngl	48986	49948	+	0
	>2245073 /158661				
60	len =	739	nex =	1	

912

	Sngl	102245	101507	-	0
	>2245073	/34167			
5	len =	1019	nex =	3	
	Init	104868	105196	+	0
	Intr	105282	105361	+	0
10	Term	105463	105866	+	0
	>2245073	/36603			
	len =	4481	nex =	11	
15	Term	6893	6584	-	0
	Intr	7287	7083	-	0
	Intr	7700	7618	-	0
	Intr	8129	7990	-	0
20	Intr	8424	8266	-	0
	Intr	9480	8479	-	0
	Intr	9839	9542	-	0
	Intr	10132	9928	-	0
	Intr	10433	10351	-	0
25	Intr	10748	10609	-	0
	Init	11064	10945	-	0
	>2245073	/37223			
30	len =	4483	nex =	11	
	Term	6893	6584	-	0
	Intr	7287	7083	-	0
	Intr	7700	7618	-	0
35	Intr	8129	7990	-	0
	Intr	8424	8266	-	0
	Intr	9480	8479	-	0
	Intr	9839	9542	-	0
	Intr	10132	9928	-	0
40	Intr	10433	10351	-	0
	Intr	10748	10609	-	0
	Init	11066	10945	-	0
	>2245073	/6042			
45	len =	959	nex =	1	
	Sngl	124096	125054	+	0
50	>2245073	/35156			
	len =	2133	nex =	7	
	Init	136139	136418	+	0
55	Intr	136654	136948	+	0
	Intr	137036	137101	+	0
	Intr	137200	137329	+	0
	Intr	137421	137579	+	0
	Intr	137703	137753	+	0
60	Term	137855	138271	+	0

>2245073 /154342

5	len =	111	nex =	1	
	Sngl	140364	140254	-	0

>2245073 /3258

10	len =	2050	nex =	8	
	Term	138586	138326	-	0
	Intr	138787	138684	-	0
	Intr	139039	138884	-	0
15	Intr	139188	139117	-	0
	Intr	139338	139291	-	0
	Intr	139469	139422	-	0
	Intr	139680	139608	-	0
	Init	140370	140183	-	0

>2245073 /2161

	len =	1690	nex =	5	
25	Init	145051	145144	+	0
	Intr	145227	145544	+	0
	Intr	145712	145798	+	0
	Intr	145888	146021	+	0
	Term	146416	146733	+	0

>2245073 /17120

	len =	464	nex =	2	
35	Init	145081	145144	+	0
	Term	145227	145544	+	0

>2245073 /29150

40	len =	1072	nex =	3	
	Init	168520	168924	+	0
	Intr	169023	169160	+	0
	Term	169230	169591	+	0

>2245073 /23025

	len =	2715	nex =	8	
50	Init	181224	181382	+	0
	Intr	181935	181992	+	0
	Intr	182407	182489	+	0
	Intr	182789	183061	+	0
	Intr	183152	183204	+	0
55	Intr	183325	183405	+	0
	Intr	183502	183614	+	0
	Term	183704	183938	+	0

>2245073 /19505

60

914

	len =	2035	nex =	4	
	Init	189969	190426	+	0
	Intr	190764	190988	+	0
5	Intr	191116	191225	+	0
	Term	191315	191480	+	0
>2245073 /31781					
10	len =	1939	nex =	4	
	Init	190050	190426	+	0
	Intr	190764	190988	+	0
	Intr	191116	191225	+	0
15	Term	191315	191480	+	0
>2245073 /36521					
20	len =	730	nex =	1	
	Sngl	190098	190332	+	0
>2245073 /39872					
25	len =	2135	nex =	4	
	Init	192291	192840	+	0
	Intr	193297	193492	+	0
	Intr	193589	193720	+	0
30	Term	194093	194425	+	0
>2245073 /6709					
35	len =	1058	nex =	2	
	Term	198909	198442	-	0
	Init	199499	199146	-	0
>2245073 /94923					
40	len =	739	nex =	2	
	Init	20607	20828	+	0
	Term	20918	21345	+	0
45	>2245073 /24997				
	len =	530	nex =	1	
50	Sngl	26357	25828	-	0
>2245073 /33509					
55	len =	1450	nex =	2	
	Init	38766	39446	+	0
	Term	39638	40214	+	0
60	>2245073 /35260				

	len =	1557	nex =	3	
	Term	43961	43535	-	0
	Intr	44176	44048	-	0
5	Init	45091	44398	-	0
	>2245073 /27500				
	len =	1700	nex =	1	
10	Sngl	51675	51471	-	0
	>2245073 /99796				
	len =	1150	nex =	2	
	Init	64024	64466	+	0
	Term	64647	65171	+	0
20	>2245073 /31538				
	len =	1278	nex =	4	
	Term	79423	79066	-	0
25	Intr	79725	79528	-	0
	Intr	80213	80047	-	0
	Init	80343	80313	-	0
	>2245073 /26448				
30	len =	2146	nex =	8	
	Term	87600	87509	-	0
	Intr	87818	87699	-	0
35	Intr	88211	88116	-	0
	Intr	88333	88295	-	0
	Intr	88636	88458	-	0
	Intr	88765	88726	-	0
	Intr	88913	88854	-	0
40	Init	89406	89167	-	0
	>2245126 /39922				
	len =	2134	nex =	5	
45	Term	28671	27817	-	0
	Intr	28825	28745	-	0
	Intr	28988	28913	-	0
	Intr	29183	29080	-	0
50	Init	29950	29830	-	0
	>2245126 /37533				
	len =	1873	nex =	7	
55	Init	30483	30887	+	0
	Intr	30977	31070	+	0
	Intr	31153	31292	+	0
	Intr	31365	31439	+	0
60	Intr	31521	31678	+	0

916

Intr	31762	31823	+	0
Term	31972	32355	+	0

>2245126 /42815

5	len =	1514	nex =	4	
	Init	56618	56988	+	0
	Intr	57254	57524	+	0
10	Intr	57621	57791	+	0
	Term	57887	58131	+	0

>2252639 /36439

15	len =	2305	nex =	12	
	Term	112752	112679	-	0
	Intr	112953	112837	-	0
	Intr	113158	113042	-	0
20	Intr	113355	113254	-	0
	Intr	113539	113444	-	0
	Intr	113704	113623	-	0
	Intr	113928	113814	-	0
	Intr	114069	114018	-	0
25	Intr	114227	114147	-	0
	Intr	114489	114328	-	0
	Intr	114748	114572	-	0
	Init	114983	114885	-	0

30 >2252639 /32628

	len =	8176	nex =	7	
	Term	112752	112549	-	0
35	Intr	112953	112837	-	0
	Intr	113158	113042	-	0
	Intr	113355	113254	-	0
	Intr	113539	113444	-	0
	Intr	113704	113623	-	0
40	Init	113928	113814	-	0

>2252639 /7870

	len =	2062	nex =	9	
45	Init	55275	55373	+	0
	Intr	55679	55864	+	0
	Intr	55943	56072	+	0
	Intr	56168	56248	+	0
50	Intr	56342	56529	+	0
	Intr	56624	56719	+	0
	Intr	56822	56915	+	0
	Intr	57043	57162	+	0
55	Term	57257	57336	+	0

>2252639 /42847

	len =	2459	nex =	3	
60	Init	64066	64204	+	0

2025 RELEASE UNDER E.O. 14176

917

	Intr	65296	65804	+	0
	Term	65895	66271	+	0
	>2252639 /20756				
5	len =	561	nex =	1	
	Sngl	66935	66375	-	0
10	>2252639 /8355				
	len =	619	nex =	1	
	Sngl	67016	66406	-	0
15	>2252639 /104398				
	len =	114	nex =	1	
20	Sngl	67655	67768	+	0
	>2252639 /34829				
	len =	1550	nex =	5	
25	Term	72152	71686	-	0
	Intr	72324	72213	-	0
	Intr	72574	72402	-	0
	Intr	72867	72664	-	0
30	Init	73235	73005	-	0
	>2252639 /34276				
	len =	2157	nex =	6	
35	Term	76139	75823	-	0
	Intr	76346	76218	-	0
	Intr	76530	76444	-	0
	Intr	76771	76626	-	0
40	Intr	76952	76898	-	0
	Init	77979	77037	-	0
	>2252639 /11108				
45	len =	539	nex =	1	
	Sngl	79342	78804	-	0
	>2252639 /1269				
50	len =	1433	nex =	3	
	Term	79851	79679	-	0
	Intr	80212	80012	-	0
55	Init	80700	80396	-	0
	>2252639 /5476				
60	len =	835	nex =	3	

ORIGINAL FILED

918

Init	85064	85271	+	0
Intr	85376	85455	+	0
Term	85554	85898	+	0

5 >2252639 /35833

len = 873 nex = 3

Init	85064	85271	+	0
Intr	85376	85455	+	0
Term	85554	85936	+	0

>2252639 /1810

15 len = 878 nex = 3

Init	85064	85271	+	0
Intr	85376	85455	+	0
Term	85554	85941	+	0

>2252639 /17857

len = 910 nex = 3

Init	85064	85271	+	0
Intr	85376	85455	+	0
Term	85554	85972	+	0

>2252639 /10862

len = 864 nex = 3

Init	85068	85271	+	0
Intr	85376	85455	+	0
Term	85554	85931	+	0

>2252639 /22773

len = 2008 nex = 2

Term	92196	90691	-	0
Init	92698	92411	-	0

>2252823 /11106

len = 1289 nex = 1

Sngl 107171 108459 + 0

50 >2252823 /25765

len = 315 nex = 1

Sngl 1671 1357 - 0

55 >2252823 /38970

len = 2486 nex = 1

60 Sngl 29968 30145 + 0

>2252823 /15741

5	len =	3070	nex =	3	
	Init	29968	30145	+	0
	Intr	30436	30547	+	0
	Term	30642	31104	+	0

10 >2252823 /28637

	len =	2900	nex =	3	
	Init	35493	36349	+	0
15	Intr	36852	37326	+	0
	Term	37673	38392	+	0

>2252823 /21038

20	len =	495	nex =	1	
	Sngl	37895	38389	+	0

25 >2252823 /35506

	len =	582	nex =	1	
	Sngl	50035	49454	-	0

30 >2252823 /39479

	len =	1604	nex =	3	
	Term	56402	56064	-	0
35	Intr	57185	56486	-	0
	Init	57649	57493	-	0

>2252823 /36326

40	len =	2392	nex =	3	
	Term	64455	64054	-	0
	Intr	64734	64625	-	0
	Init	65205	64824	-	0

45 >2252823 /31027

	len =	1150	nex =	2	
50	Init	94085	94153	+	0
	Term	94219	95230	+	0

>2252848 /111719

55	len =	733	nex =	1	
	Sngl	46064	45332	-	0

60 >2252848 /11036

920

	len =	790	nex =	1	
	Sngl	46089	45304	-	0
5	>2252848	/3204			
	len =	833	nex =	1	
	Sngl	60597	61429	+	0
10	>2252848	/22161			
	len =	670	nex =	1	
15	Sngl	63070	63731	+	0
	>2252848	/22348			
	len =	740	nex =	1	
20	Sngl	65608	64869	-	0
	>2252848	/28082			
25	len =	1216	nex =	3	
	Init	80915	80991	+	0
	Intr	81337	81552	+	0
30	Term	81645	81897	+	0
	>2252848	/26442			
	len =	1210	nex =	3	
35	Init	80915	80991	+	0
	Intr	81337	81552	+	0
	Term	81645	81895	+	0
	>2252848	/37305			
40	len =	1575	nex =	4	
	Term	91905	91570	-	0
	Intr	92168	92002	-	0
45	Intr	92528	92246	-	0
	Init	92758	92613	-	0
	>2252848	/37175			
50	len =	2050	nex =	3	
	Term	95449	94674	-	0
	Intr	95668	95551	-	0
	Init	96720	96101	-	0
55	>2262097	/22611			
	len =	1439	nex =	4	
60	Init	31	168	+	0

921

Intr	253	403	+	0
Intr	481	885	+	0
Term	969	1469	+	0

5 >2262097 /37663

len = 1694 nex = 2

Term	48814	47723	-	0
Init	49413	49234	-	0

>2262097 /37704

15 len = 1990 nex = 6

Term	4521	4199	-	0
Intr	4778	4665	-	0
Intr	5379	5207	-	0
Intr	5540	5489	-	0
Intr	5680	5632	-	0
Init	6186	5782	-	0

>2262097 /112955

25 len = 2350 nex = 4

Term	89371	88825	-	0
Intr	89563	89456	-	0
Intr	89803	89654	-	0
Init	91172	90509	-	0

>2262135 /41490

35 len = 1454 nex = 2

Term	2318	1916	-	0
Init	3369	2625	-	0

>2262135 /20167

40 len = 1304 nex = 2

Term	4241	3765	-	0
Init	5068	4768	-	0

45 len = 1390 nex = 3

>2262135 /32291

len = 1390 nex = 3

Term	3887	3685	-	0
Intr	4241	4100	-	0
Init	5072	4768	-	0

>2262135 /6568

55 len = 2212 nex = 6

Term	55501	55152	-	0
Intr	55716	55591	-	0
Intr	55868	55793	-	0

60 len = 2212 nex = 6

20250322:001600

922

	Intr	56088	55950	-	0
	Intr	56564	56483	-	0
	Init	57080	56653	-	0
5	>2262135 /10207				
	len =	2063	nex =	4	
10	Init	59951	60024	+	0
	Intr	60681	60762	+	0
	Intr	61016	61098	+	0
	Term	61517	61813	+	0
15	>2262135 /18545				
	len =	647	nex =	1	
	Sngl	6145	6791	+	0
20	>2262135 /4346				
	len =	2939	nex =	6	
25	Init	70603	71150	+	0
	Intr	71555	71677	+	0
	Intr	71842	71907	+	0
	Intr	71994	72059	+	0
	Intr	72734	72814	+	0
	Term	72893	73541	+	0
30	>2262135 /26127				
	len =	817	nex =	1	
35	Sngl	10051	10199	+	0
	>2262135 /8114				
40	len =	1879	nex =	4	
	Init	97068	97416	+	0
	Intr	98158	98297	+	0
	Intr	98468	98540	+	0
	Term	98650	98946	+	0
45	>2262135 /34186				
	len =	347	nex =	1	
50	Sngl	97069	97415	+	0
	>2262135 /145375				
55	len =	319	nex =	1	
	Sngl	10051	10164	+	0
	>2262135 /18454				
60	len =	354	nex =	1	

2025-09-23 09:00:00

	Sngl	10051	10199	+	0
	>2262135	/27915			
5	len =	1173	nex =	3	
	Init	99470	99712	+	0
	Intr	99822	99870	+	0
10	Term	99982	100642	+	0
	>2262155	/1441			
	len =	657	nex =	1	
15	Sngl	23119	22463	-	0
	>2262155	/38365			
20	len =	2443	nex =	11	
	Term	33741	33609	-	0
	Intr	33874	33812	-	0
	Intr	34038	33961	-	0
25	Intr	34207	34130	-	0
	Intr	34357	34283	-	0
	Intr	34542	34456	-	0
	Intr	35004	34864	-	0
	Intr	35174	35106	-	0
30	Intr	35320	35254	-	0
	Intr	35536	35471	-	0
	Init	36051	35849	-	0
	>2262155	/2578			
35	len =	2710	nex =	12	
	Term	41819	41536	-	0
	Intr	42007	41945	-	0
40	Intr	42177	42100	-	0
	Intr	42353	42276	-	0
	Intr	42507	42433	-	0
	Intr	42691	42605	-	0
	Intr	42920	42792	-	0
45	Intr	43144	43004	-	0
	Intr	43300	43232	-	0
	Intr	43448	43382	-	0
	Intr	43690	43625	-	0
	Init	44238	44044	-	0
50	>2262155	/10042			
	len =	1776	nex =	4	
55	Init	47118	47195	+	0
	Intr	47279	47459	+	0
	Intr	47575	47672	+	0
	Term	47837	48384	+	0
60	>2262155	/13246			

2025 RELEASE UNDER E.O. 14176

	len =	1990	nex =	6	
	Init	54079	54165	+	0
5	Intr	54255	54346	+	0
	Intr	54432	54540	+	0
	Intr	54640	54675	+	0
	Intr	54764	54850	+	0
	Term	54940	55113	+	0
10	>2262155 /34698				
	len =	1459	nex =	6	
15	Init	56211	56260	+	0
	Intr	56344	56556	+	0
	Intr	56654	56802	+	0
	Intr	56878	57034	+	0
	Intr	57160	57252	+	0
20	Term	57530	57669	+	0
	>2262155 /39211				
	len =	2110	nex =	2	
25	Init	64477	65546	+	0
	Term	66273	66579	+	0
	>2262155 /19601				
30	len =	2050	nex =	2	
	Init	64534	65546	+	0
	Term	66273	66579	+	0
35	>2262155 /32751				
	len =	850	nex =	1	
40	Sngl	77445	76604	-	0
	>2262155 /3276				
	len =	1167	nex =	1	
45	Sngl	8628	9794	+	0
	>2264302 /38370				
50	len =	1450	nex =	2	
	Term	35101	34004	-	0
	Init	35452	35188	-	0
55	>2264302 /9562				
	len =	2074	nex =	0	
60	>2264302 /28046				

925

	len =	1581	nex =	4	
	Term	51719	51257	-	0
	Intr	52040	51910	-	0
5	Intr	52474	52402	-	0
	Init	52837	52724	-	0
	>2264302 /16428				
10	len =	1571	nex =	4	
	Term	51818	51294	-	0
	Intr	52040	51910	-	0
	Intr	52474	52402	-	0
15	Init	52864	52724	-	0
	>2264302 /100085				
20	len =	1254	nex =	3	
	Term	5287	4881	-	0
	Intr	5613	5357	-	0
	Init	6134	5782	-	0
25	>2264303 /22				
	len =	1735	nex =	6	
	Init	14289	14642	+	0
30	Intr	14799	14910	+	0
	Intr	15002	15095	+	0
	Intr	15228	15405	+	0
	Intr	15488	15557	+	0
35	Term	15638	16023	+	0
	>2264303 /7145				
	len =	824	nex =	4	
40	Init	3387	3465	+	0
	Intr	3544	3666	+	0
	Intr	3754	3870	+	0
	Term	3947	4205	+	0
45	>2264303 /4273				
	len =	1845	nex =	3	
	Term	45044	44650	-	0
50	Intr	45266	45126	-	0
	Init	46494	46178	-	0
	>2264303 /35612				
55	len =	1469	nex =	4	
	Init	58748	59002	+	0
	Intr	59229	59277	+	0
	Intr	59634	59833	+	0
60	Term	59930	60216	+	0

>2264303 /42336

5	len =	1825	nex =	4	
	Term	64023	63682	-	0
	Intr	64570	64473	-	0
	Intr	65089	64989	-	0
	Init	65506	65289	-	0

10 >2264304 /34402

	len =	2558	nex =	5	
15	Init	20281	20902	+	0
	Intr	21285	21510	+	0
	Intr	21627	21849	+	0
	Intr	22104	22317	+	0
	Term	22554	22838	+	0

20 >2264304 /34783

	len =	2075	nex =	5	
25	Term	23983	23714	-	0
	Intr	24174	24080	-	0
	Intr	24709	24267	-	0
	Intr	25149	24793	-	0
	Init	25788	25400	-	0

30 >2264304 /39319

	len =	1870	nex =	5	
35	Init	2871	2989	+	0
	Intr	3690	3771	+	0
	Intr	3960	4165	+	0
	Intr	4328	4381	+	0
	Term	4476	4733	+	0

40 >2264304 /9159

	len =	1570	nex =	2	
45	Init	41803	42064	+	0
	Term	42974	43372	+	0

>2264304 /38464

50	len =	1270	nex =	1	
	Sngl	51034	52303	+	0

>2264304 /28578

55	len =	2110	nex =	5	
	Init	515	1139	+	0
	Intr	1407	1504	+	0
60	Intr	1754	1853	+	0

2025 RELEASE UNDER E.O. 14176

927

	Intr	2027	2272	+	0
	Term	2358	2618	+	0
	>2264304 /41195				
5	len =	353	nex =	1	
	Sngl	57898	57549	-	0
10	>2264304 /2871				
	len =	430	nex =	2	
	Init	6595	6647	+	0
15	Term	6733	7019	+	0
	>2264304 /30073				
	len =	1810	nex =	1	
20	Sngl	65320	65030	-	0
	>2264304 /32071				
25	len =	1128	nex =	1	
	Sngl	67814	67283	-	0
	>2264304 /103464				
30	len =	1096	nex =	1	
	Sngl	67814	67316	-	0
35	>2264304 /17818				
	len =	1136	nex =	1	
40	Sngl	67814	67277	-	0
	>2264304 /24095				
	len =	596	nex =	1	
45	Sngl	72223	72818	+	0
	>2264304 /111741				
	len =	2898	nex =	9	
50	Init	77610	77692	+	0
	Intr	78044	78153	+	0
	Intr	78600	78734	+	0
	Intr	78876	79022	+	0
55	Intr	79400	79483	+	0
	Intr	79589	79635	+	0
	Intr	79729	79802	+	0
	Intr	79915	79973	+	0
60	Term	80152	80212	+	0

>2264305 /10263

len =	1493	nex =	5	
5	Init	31119	31386	+
	Intr	31604	31784	+
	Intr	31864	32005	+
	Intr	32090	32159	+
	Term	32249	32611	+

10 >2264305 /98400

len =	993	nex =	3	
15	Term	4415	4173	-
	Intr	4868	4742	-
	Init	5152	4965	-

20 >2264305 /36333

len =	1450	nex =	4	
	Term	4415	4119	-
	Intr	4868	4742	-
25	Intr	5244	4965	-
	Init	5422	5374	-

>2264305 /121728

30	len =	550	nex =	2	
	Term	5244	5080	-	0
	Init	5422	5374	-	0

35 >2264305 /41072

len =	1312	nex =	4	
	Term	4415	4326	-
40	Intr	4868	4742	-
	Intr	5244	4965	-
	Init	5422	5374	-

>2264305 /24983

45	len =	599	nex =	1	
	Sngl	64677	64079	-	0

50 >2264305 /16865

len =	1615	nex =	4	
	Init	71009	71096	+
55	Intr	71447	71574	+
	Intr	71737	71841	+
	Term	72035	72347	+

>2264305 /35698

60

2025 RELEASE UNDER E.O. 14176

	len =	1150	nex =	4	
	Init	71025	71096	+	0
	Intr	71447	71574	+	0
5	Intr	71737	71841	+	0
	Term	72035	72162	+	0
	>2264306 /21505				
10	len =	1450	nex =	3	
	Term	10517	10132	-	0
	Intr	11048	10721	-	0
	Init	11577	11269	-	0
15	>2264306 /19024				
	len =	715	nex =	2	
20	Term	14439	14066	-	0
	Init	14777	14527	-	0
	>2264306 /33140				
25	len =	1450	nex =	3	
	Term	14439	13966	-	0
	Intr	14854	14527	-	0
	Init	15411	14979	-	0
30	>2264306 /121213				
	len =	333	nex =	1	
35	Sngl	2596	2928	+	0
	>2264306 /39888				
40	len =	2203	nex =	9	
	Term	35099	34644	-	0
	Intr	35279	35181	-	0
	Intr	35475	35371	-	0
	Intr	35651	35559	-	0
45	Intr	35855	35763	-	0
	Intr	36011	35958	-	0
	Intr	36218	36117	-	0
	Intr	36369	36295	-	0
	Init	36846	36503	-	0
50	>2264306 /11054				
	len =	1417	nex =	5	
55	Init	41110	41228	+	0
	Intr	41333	41424	+	0
	Intr	41763	41818	+	0
	Intr	42120	42181	+	0
	Term	42324	42526	+	0
60					

20250826 09:23:50

930

>2264306 /3699

```
len = 1897    nex = 4
```

5	Init	5030	5266	+	0
	Intr	5420	6238	+	0
	Intr	6325	6526	+	0
	Term	6551	6926	+	0

10 >2264306 /6637

```
len = 1428    nex = 3
```

	Term	80382	79690	-	0
15	Intr	80764	80484	-	0
	Init	81117	80852	-	0

>2264306 /111669

```
20      len =      382      nex =      2
```

Init	88535	88581	+	0
Term	88664	88916	+	0

25 >2264307 /42441

```
len =      682    nex =      2
```

	Term	48650	48344	-	0
30	Init	49017	48966	-	0

>2264307 /22848

```
len = 658    nex = 2
```

Term	48650	48368	-	0
Init	49017	48966	-	0

>2264307 /145394

```
len =      638      nex =      2
```

Term	48650	48388	-	0
Init	49017	48966	-	0

45 >2264307 /11511

```
len =      776      nex =      2
```

50	Term	48650	48252	-	0
	Init	49027	48966	-	0

>2264307 /12330

```
55      len =      670      nex =      2
```

Term	48650	48363	-	0
Init	49017	48966	-	0

60 >2264307 /37668

	len =	2959	nex =	12	
	Term	58676	58435	-	0
5	Intr	58819	58762	-	0
	Intr	59006	58939	-	0
	Intr	59148	59089	-	0
	Intr	59415	59374	-	0
	Intr	59547	59504	-	0
10	Intr	59753	59684	-	0
	Intr	60223	60104	-	0
	Intr	60499	60481	-	0
	Intr	60688	60616	-	0
	Intr	60911	60847	-	0
15	Init	61393	61056	-	0
	>2264307	/24058			
	len =	1653	nex =	4	
20	Init	72492	72816	+	0
	Intr	73287	73411	+	0
	Intr	73485	73593	+	0
	Term	73888	74144	+	0
25	>2264308	/1935			
	len =	1396	nex =	1	
30	Sngl	17599	16204	-	0
	>2264308	/22483			
	len =	2981	nex =	8	
35	Term	4792	4416	-	0
	Intr	5296	4866	-	0
	Intr	5495	5375	-	0
	Intr	5737	5588	-	0
40	Intr	6028	5823	-	0
	Intr	6224	6110	-	0
	Intr	6544	6307	-	0
	Init	7396	7131	-	0
45	>2264309	/37959			
	len =	357	nex =	1	
	Sngl	16800	16444	-	0
50	>2264309	/15155			
	len =	872	nex =	2	
55	Init	22581	22830	+	0
	Term	22927	23337	+	0
	>2264309	/36334			
60	len =	4030	nex =	8	

>2264310 /17408

len = 642 nex = 1

5 Sngl 75188 74547 - 0

>2264310 /125083

len = 1961 nex = 5

10 Init 8184 8440 + 0

Intr 8574 8786 + 0

Intr 8879 9037 + 0

Intr 9616 9684 + 0

15 Term 9797 10144 + 0

>2264311 /32868

len = 1724 nex = 5

20 Term 22845 22268 - 0

Intr 23036 22924 - 0

Intr 23230 23115 - 0

25 Intr 23684 23307 - 0

Init 23977 23868 - 0

>2264311 /6256

len = 970 nex = 3

30 Term 61688 61655 - 0

Intr 61915 61777 - 0

Init 62223 62000 - 0

35 >2264311 /125951

len = 2213 nex = 8

40 Term 60708 60456 - 0

Intr 60920 60814 - 0

Intr 61074 61009 - 0

Intr 61491 61410 - 0

Intr 61688 61644 - 0

Intr 61915 61777 - 0

45 Intr 62223 62000 - 0

Init 62668 62430 - 0

>2264311 /27195

50 len = 1880 nex = 7

Term 82920 82401 - 0

Intr 83150 83009 - 0

Intr 83482 83243 - 0

55 Intr 83616 83581 - 0

Intr 83788 83708 - 0

Intr 83928 83871 - 0

Init 84280 84011 - 0

60 >2264312 /14950

934

	len =	881	nex =	1	
5	Sngl	27808	26928	-	0
	>2264312	/95433			
	len =	1318	nex =	5	
10	Term	41828	41661	-	0
	Intr	42031	41958	-	0
	Intr	42285	42119	-	0
	Intr	42519	42450	-	0
	Init	42741	42601	-	0
15	>2264312	/41937			
	len =	412	nex =	1	
20	Sngl	46315	45915	-	0
	>2264312	/13715			
	len =	505	nex =	1	
25	Sngl	46419	45915	-	0
	>2264312	/20908			
30	len =	1588	nex =	1	
	Sngl	47047	45915	-	0
	>2264312	/121153			
35	len =	1599	nex =	0	
	>2264312	/21872			
40	len =	1999	nex =	5	
	Init	76178	76439	+	0
	Intr	76875	77278	+	0
	Intr	77349	77609	+	0
45	Intr	77680	77802	+	0
	Term	77884	78176	+	0
	>2264312	/40252			
50	len =	929	nex =	3	
	Init	8129	8281	+	0
	Intr	8374	8529	+	0
	Term	8834	9057	+	0
55	>2264313	/13012			
	len =	2530	nex =	3	
60	Init	50735	51416	+	0

935

Intr	51723	52053	+	0
Term	52969	53262	+	0

>2264313 /156373

5	len =	1597	nex =	4
---	-------	------	-------	---

Term	56197	55946	-	0	
Intr	56442	56319	-	0	
10	Intr	57210	56988	-	0
	Init	57542	57464	-	0

>2264314 /8635

15	len =	1886	nex =	4
----	-------	------	-------	---

Term	10067	9103	-	0	
Intr	10250	10148	-	0	
20	Intr	10433	10340	-	0
	Init	10988	10835	-	0

>2264314 /115644

25	len =	1259	nex =	2
----	-------	------	-------	---

Term	26540	26126	-	0
Init	27384	26837	-	0

>2264314 /38996

30	len =	2313	nex =	7
----	-------	------	-------	---

Term	27833	27526	-	0	
Intr	28049	27984	-	0	
35	Intr	28349	28278	-	0
	Intr	28813	28492	-	0
	Intr	29046	28886	-	0
	Intr	29175	29131	-	0
40	Init	29838	29580	-	0

>2264314 /32785

len =	1499	nex =	1
-------	------	-------	---

45	Sngl	41738	42167	+	0
----	------	-------	-------	---	---

>2264314 /20245

len =	1429	nex =	1
-------	------	-------	---

50	Sngl	41738	42147	+	0
----	------	-------	-------	---	---

>2264314 /5592

55	len =	1450	nex =	0
----	-------	------	-------	---

>2264314 /13819

60	len =	1390	nex =	1
----	-------	------	-------	---

936

	Sngl	41738	42167	+	0
	>2264314	/29726			
5	len =	673	nex =	1	
	Sngl	46055	46727	+	0
	>2264314	/41900			
10	len =	567	nex =	1	
	Sngl	46131	46697	+	0
15	>2264314	/2462			
	len =	570	nex =	1	
	Sngl	46131	46700	+	0
20	>2264314	/16750			
	len =	585	nex =	1	
25	Sngl	46131	46715	+	0
	>2264314	/18232			
	len =	1571	nex =	5	
30	Term	48315	47879	-	0
	Intr	48456	48413	-	0
	Intr	48598	48541	-	0
	Intr	48919	48826	-	0
35	Init	49449	49182	-	0
	>2264314	/9012			
	len =	1870	nex =	0	
40	>2264314	/7365			
	len =	1776	nex =	0	
45	>2264314	/33059			
	len =	2811	nex =	7	
	Term	61633	61320	-	0
50	Intr	61973	61823	-	0
	Intr	62227	62054	-	0
	Intr	62409	62320	-	0
	Intr	62646	62576	-	0
	Intr	63811	62772	-	0
55	Init	64130	63836	-	0
	>2264314	/27647			
60	len =	1370	nex =	3	

937

	Init	72212	72591	+	0
	Intr	72849	73086	+	0
	Term	73196	73581	+	0
5	>2264315	/10218			
	len =	2270	nex =	4	
	Term	26015	25438	-	0
10	Intr	26141	26094	-	0
	Intr	27175	26240	-	0
	Init	27707	27384	-	0
	>2264315	/29462			
15	len =	1139	nex =	2	
	Init	45117	45873	+	0
	Term	45961	46255	+	0
20	>2264315	/14965			
	len =	430	nex =	1	
	Sngl	47036	46610	-	0
25	>2264315	/114307			
	len =	464	nex =	1	
30	Sngl	47105	46642	-	0
	>2264315	/3363			
	len =	636	nex =	1	
35	Sngl	47111	46476	-	0
	>2264315	/41666			
40	len =	2157	nex =	12	
	Init	59476	59703	+	0
	Intr	59800	59887	+	0
	Intr	60015	60074	+	0
45	Intr	60160	60192	+	0
	Intr	60278	60355	+	0
	Intr	60433	60476	+	0
	Intr	60582	60622	+	0
	Intr	60709	60791	+	0
50	Intr	60876	60967	+	0
	Intr	61055	61124	+	0
	Intr	61205	61246	+	0
	Term	61348	61632	+	0
55	>2264316	/31759			
	len =	1810	nex =	3	
	Term	40887	40024	-	0
60	Intr	41245	40976	-	0

938

	Init	41826	41375	-	0
	>2264316 /4716				
5	len =	1150	nex =	4	
	Term	48078	47771	-	0
	Intr	48347	48169	-	0
	Intr	48549	48448	-	0
10	Init	48918	48760	-	0
	>2264316 /35357				
	len =	3430	nex =	2	
15	Init	4937	5508	+	0
	Term	7116	8360	+	0
	>2264316 /13418				
20	len =	1121	nex =	4	
	Term	50134	49841	-	0
	Intr	50452	50271	-	0
25	Intr	50665	50567	-	0
	Init	50961	50832	-	0
	>2264316 /25839				
30	len =	1733	nex =	4	
	Term	52037	51717	-	0
	Intr	52799	52621	-	0
	Intr	52994	52893	-	0
35	Init	53449	53248	-	0
	>2264316 /5103				
	len =	566	nex =	2	
40	Term	56108	55749	-	0
	Init	56314	56188	-	0
	>2264316 /25723				
45	len =	1118	nex =	3	
	Init	70502	70609	+	0
	Intr	70687	70765	+	0
50	Term	71265	71619	+	0
	>2264316 /28686				
	len =	1761	nex =	5	
55	Init	73159	73478	+	0
	Intr	73823	73864	+	0
	Intr	74151	74238	+	0
	Intr	74355	74436	+	0
60	Term	74532	74919	+	0

>2264316 /33187

5	len =	1316	nex =	6		
	Init	75294	75411	+	0	
	Intr	75493	75533	+	0	
	Intr	75623	75723	+	0	
10	Intr	75977	76121	+	0	
	Intr	76215	76304	+	0	
	Term	76389	76609	+	0	

>2264316 /40559

15	len =	940	nex =	4		
	Init	75623	75723	+	0	
	Intr	75977	76121	+	0	
	Intr	76215	76304	+	0	
20	Term	76389	76430	+	0	

>2264317 /27304

25	len =	1450	nex =	4		
	Init	10536	10865	+	0	
	Intr	11094	11307	+	0	
	Intr	11430	11575	+	0	
30	Term	11678	11977	+	0	

>2264317 /41386

35	len =	2230	nex =	7		
	Init	18624	18806	+	0	
	Intr	19320	19433	+	0	
	Intr	19544	19688	+	0	
	Intr	19786	19863	+	0	
	Intr	19964	20076	+	0	
40	Intr	20166	20269	+	0	
	Term	20364	20848	+	0	

>2264317 /19638

45	len =	1116	nex =	5		
	Term	39626	39380	-	0	
	Intr	39837	39741	-	0	
	Intr	39994	39932	-	0	
50	Intr	40263	40110	-	0	
	Init	40495	40353	-	0	

>2264317 /6734

55	len =	2230	nex =	8		
	Init	43041	43121	+	0	
	Intr	43615	43732	+	0	
	Intr	43820	43927	+	0	
60	Intr	44029	44153	+	0	

2025 RELEASE UNDER E.O. 14176

940

	Intr	44256	44520	+	0
	Intr	44612	44680	+	0
	Intr	44773	44934	+	0
	Term	45031	45269	+	0
5	>2264318 /3797				
	len =	3010	nex =	11	
10	Term	14549	14209	-	0
	Intr	14698	14642	-	0
	Intr	14911	14777	-	0
	Intr	15084	15004	-	0
	Intr	15230	15162	-	0
15	Intr	15408	15334	-	0
	Intr	15837	15757	-	0
	Intr	16050	15932	-	0
	Intr	16304	16139	-	0
	Intr	16522	16393	-	0
20	Init	17210	16609	-	0
	>2264318 /33231				
	len =	1510	nex =	4	
25	Term	19006	18683	-	0
	Intr	19387	19102	-	0
	Intr	19635	19485	-	0
	Init	20191	19835	-	0
30	>2264318 /42276				
	len =	681	nex =	1	
35	Sngl	24794	24114	-	0
	>2264318 /26752				
	len =	754	nex =	1	
40	Sngl	6372	6627	+	0
	>2264318 /25855				
45	len =	2410	nex =	4	
	Init	74093	74435	+	0
	Intr	74770	74907	+	0
	Intr	75288	75359	+	0
50	Term	75730	76502	+	0
	>2264319 /37985				
	len =	1041	nex =	3	
55	Term	29497	28961	-	0
	Intr	29867	29820	-	0
	Init	30001	29945	-	0
60	>2264320 /36697				

	len =	3070	nex =	12	
	Init	77774	77917	+	0
5	Intr	78577	78735	+	0
	Intr	78827	78886	+	0
	Intr	79001	79047	+	0
	Intr	79159	79212	+	0
	Intr	79302	79479	+	0
10	Intr	79602	79754	+	0
	Intr	79848	79913	+	0
	Intr	80000	80047	+	0
	Intr	80127	80233	+	0
	Intr	80327	80405	+	0
15	Term	80513	80843	+	0
>2264321 /22350					
	len =	1378	nex =	1	
20	Sngl	42485	43862	+	0
>2264321 /17195					
	len =	2639	nex =	9	
	Term	44138	43885	-	0
	Intr	44545	44459	-	0
	Intr	44723	44638	-	0
30	Intr	45000	44910	-	0
	Intr	45164	45079	-	0
	Intr	45369	45261	-	0
	Intr	45565	45519	-	0
	Intr	45725	45656	-	0
35	Init	46523	45984	-	0
>2264321 /4025					
	len =	1845	nex =	8	
40	Term	62897	62608	-	0
	Intr	63027	62981	-	0
	Intr	63329	63108	-	0
	Intr	63461	63409	-	0
45	Intr	63720	63555	-	0
	Intr	63853	63812	-	0
	Intr	64024	63940	-	0
	Init	64452	64119	-	0
50	>2264321 /226				
	len =	1214	nex =	3	
	Init	64562	64824	+	0
55	Intr	65227	65327	+	0
	Term	65506	65775	+	0
>2264321 /25843					
60	len =	1179	nex =	1	

20250726 09:16:00

942

	Sngl	64602	65780	+	0
5	>2264367	/13226			
	len =	760	nex =	1	
	Sngl	17702	16945	-	0
10	>2264367	/6280			
	len =	1721	nex =	4	
15	Init	79635	80401	+	0
	Intr	80649	80739	+	0
	Intr	80875	81047	+	0
	Term	81136	81355	+	0
20	>2264367	/14253			
	len =	394	nex =	1	
	Sngl	79694	80087	+	0
25	>2264367	/2093			
	len =	1697	nex =	4	
30	Term	81924	81450	-	0
	Intr	82092	82014	-	0
	Intr	82411	82172	-	0
	Init	83146	82545	-	0
35	>2275194	/35109			
	len =	1541	nex =	0	
	>2275194	/20378			
40	len =	540	nex =	1	
	Sngl	46427	45888	-	0
45	>2275194	/6324			
	len =	564	nex =	1	
	Sngl	81129	81692	+	0
50	>2275194	/95662			
	len =	550	nex =	0	
55	>2275194	/21715			
	len =	339	nex =	1	
	Sngl	81340	81678	+	0
60	>2275194	/34414			

943

	len =	2177	nex =	6	
5	Init	1529	1687	+	0
	Intr	1807	1877	+	0
	Intr	2195	2314	+	0
	Intr	2406	2524	+	0
	Intr	2616	2697	+	0
	Term	2789	3076	+	0
10	>2275194 /35584				
	len =	2064	nex =	6	
15	Init	1529	1687	+	0
	Intr	1807	1877	+	0
	Intr	2195	2314	+	0
	Intr	2406	2524	+	0
	Intr	2616	2697	+	0
	Term	2789	3017	+	0
20	>2281081 /99937				
	len =	1179	nex =	4	
25	Init	17994	18277	+	0
	Intr	18570	18617	+	0
	Intr	18757	18836	+	0
	Term	18973	19172	+	0
30	>2281081 /34407				
	len =	1614	nex =	2	
35	Term	20892	20042	-	0
	Init	21655	20980	-	0
	>2281081 /24415				
40	len =	1398	nex =	1	
	Sngl	37217	37695	+	0
	>2281081 /21866				
45	len =	3043	nex =	10	
	Init	41405	41802	+	0
	Intr	41989	42098	+	0
50	Intr	42186	42243	+	0
	Intr	42347	42610	+	0
	Intr	42881	43018	+	0
	Intr	43151	43202	+	0
	Intr	43288	43367	+	0
55	Intr	43475	43534	+	0
	Intr	43663	43743	+	0
	Term	44186	44447	+	0
60	>2281081 /117763				

944

	len =	168	nex =	1	
	Sngl	44280	44447	+	0
5	>2281081	/37969			
	len =	1630	nex =	2	
10	Init	45636	46252	+	0
	Term	46437	47256	+	0
	>2281081	/97249			
15	len =	1570	nex =	3	
	Init	75474	75567	+	0
	Intr	75664	75773	+	0
	Term	76110	76381	+	0
20	>2288979	/30737			
	len =	1957	nex =	6	
25	Term	23749	23549	-	0
	Intr	24382	24215	-	0
	Intr	24583	24465	-	0
	Intr	24734	24673	-	0
	Intr	24906	24830	-	0
	Init	25505	25278	-	0
30	>2288979	/42038			
	len =	2417	nex =	6	
35	Term	26123	25700	-	0
	Intr	26352	26213	-	0
	Intr	26728	26523	-	0
	Intr	27113	27007	-	0
	Intr	27509	27330	-	0
40	Init	28116	27832	-	0
	>2288979	/5460			
	len =	1369	nex =	2	
45	Term	61213	60939	-	0
	Init	61831	61648	-	0
	>2288979	/31535			
50	len =	971	nex =	1	
	Sngl	6953	5983	-	0
55	>2288979	/15927			
	len =	582	nex =	2	
60	Init	83467	83577	+	0
	Term	83732	84048	+	0

	len =	510	nex =	2	
	Term	89987	89866	-	0
	Init	90375	90255	-	0
5	>2326340 /17730				
	len =	938	nex =	3	
10	Init	12848	12929	+	0
	Intr	13222	13294	+	0
	Term	13456	13785	+	0
	>2335089 /17415				
15	len =	1711	nex =	2	
	Init	18997	19833	+	0
	Term	20359	20707	+	0
20	>2335089 /41462				
	len =	2561	nex =	7	
25	Init	77553	77859	+	0
	Intr	78200	78282	+	0
	Intr	78527	78615	+	0
	Intr	78796	78869	+	0
	Intr	78950	79000	+	0
30	Intr	79347	79408	+	0
	Term	79492	80113	+	0
	>2337888 /30632				
35	len =	597	nex =	1	
	Sngl	45399	44803	-	0
	>2337888 /33132				
40	len =	1427	nex =	1	
	Sngl	56372	54946	-	0
45	>2337888 /25271				
	len =	1190	nex =	4	
	Term	81979	81460	-	0
50	Intr	82251	82069	-	0
	Intr	82443	82348	-	0
	Init	82649	82529	-	0
	>2337888 /36364				
55	len =	2473	nex =	10	
	Term	81979	81474	-	0
	Intr	82251	82069	-	0
60	Intr	82443	82348	-	0

947

	Intr	82588	82529	-	0
	Intr	82726	82673	-	0
	Intr	82906	82829	-	0
	Intr	83042	82989	-	0
5	Intr	83230	83147	-	0
	Intr	83655	83627	-	0
	Init	83946	83768	-	0
	>2337888 /48				
10	len =	139	nex =	1	
	Sngl	84105	83967	-	0
	>2337888 /39291				
15	len =	1330	nex =	2	
	Init	9724	10277	+	0
20	Term	10380	11048	+	0
	>2341023 /20848				
25	len =	2394	nex =	8	
	Term	105381	105134	-	0
	Intr	105770	105531	-	0
	Intr	106011	105948	-	0
	Intr	106356	106242	-	0
30	Intr	106669	106531	-	0
	Intr	106971	106841	-	0
	Intr	107209	107080	-	0
	Init	107527	107476	-	0
35	>2341023 /4513				
	len =	1287	nex =	3	
	Term	16071	15822	-	0
40	Intr	16960	16676	-	0
	Init	17108	17082	-	0
	>2341023 /26558				
45	len =	1150	nex =	3	
	Term	23857	23331	-	0
	Intr	24045	23945	-	0
	Init	24472	24392	-	0
50	>2341023 /23398				
	len =	2892	nex =	2	
55	Term	36567	36137	-	0
	Init	39028	38927	-	0
	>2341023 /40467				
60	len =	2202	nex =	7	

20250919-001600

948

	Init	41815	41979	+	0
	Intr	42299	42457	+	0
	Intr	42564	42739	+	0
5	Intr	42897	43174	+	0
	Intr	43264	43399	+	0
	Intr	43492	43603	+	0
	Term	43692	44016	+	0
10	>2341023		/19832		
	len =	2656	nex =	4	
	Init	45198	45615	+	0
15	Intr	45720	45944	+	0
	Intr	46040	46752	+	0
	Term	46898	47342	+	0
20	>2341023		/91880		
	len =	1118	nex =	2	
	Init	46306	46752	+	0
	Term	46898	47423	+	0
25	>2341023		/8374		
	len =	805	nex =	3	
30	Init	84788	85031	+	0
	Intr	85113	85256	+	0
	Term	85340	85592	+	0
35	>2341023		/9471		
	len =	649	nex =	1	
	Sngl	85423	85236	-	0
40	>2341023		/30909		
	len =	1020	nex =	3	
	Init	90351	90483	+	0
45	Intr	90571	90628	+	0
	Term	91104	91353	+	0
	>2341023		/28606		
50	len =	730	nex =	1	
	Sngl	91839	92568	+	0
55	>2341023		/125151		
	len =	310	nex =	1	
	Sngl	96904	96600	-	0
60	>2341023		/33613		

	len =	2290	nex =	5	
	Term	94901	94658	-	0
5	Intr	95464	95403	-	0
	Intr	95744	95606	-	0
	Intr	96270	96059	-	0
	Init	96946	96584	-	0
10	>2342673	/21644			
	len =	568	nex =	2	
	Init	1	19	+	0
15	Term	287	568	+	0
	>2342673	/4236			
	len =	1031	nex =	1	
20	Sngl	15499	14469	-	0
	>2342673	/13218			
25	len =	600	nex =	1	
	Sngl	59777	59178	-	0
	>2342673	/1911			
30	len =	1410	nex =	0	
	>2342673	/15745			
35	len =	2693	nex =	15	
	Term	72951	72598	-	0
	Intr	73173	73059	-	0
	Intr	73327	73268	-	0
40	Intr	73473	73420	-	0
	Intr	73651	73592	-	0
	Intr	73809	73747	-	0
	Intr	73936	73893	-	0
	Intr	74109	74025	-	0
45	Intr	74283	74203	-	0
	Intr	74471	74379	-	0
	Intr	74618	74554	-	0
	Intr	74789	74714	-	0
	Intr	74956	74891	-	0
50	Intr	75176	75051	-	0
	Init	75290	75255	-	0
	>2342673	/20814			
55	len =	2669	nex =	14	
	Term	87698	87414	-	0
	Intr	87906	87792	-	0
	Intr	88057	87998	-	0
60	Intr	88219	88166	-	0

950

	Intr	88375	88316	-	0
	Intr	88529	88467	-	0
	Intr	88664	88621	-	0
5	Intr	88853	88769	-	0
	Intr	89044	88964	-	0
	Intr	89241	89149	-	0
	Intr	89408	89344	-	0
	Intr	89583	89508	-	0
	Intr	89751	89686	-	0
10	Init	89916	89851	-	0

>2342673 /36585

	len =	3206	nex =	16	
15	Term	87698	87522	-	0
	Intr	87906	87792	-	0
	Intr	88057	87998	-	0
20	Intr	88219	88166	-	0
	Intr	88375	88316	-	0
	Intr	88529	88467	-	0
	Intr	88664	88621	-	0
	Intr	88853	88769	-	0
25	Intr	89044	88964	-	0
	Intr	89241	89149	-	0
	Intr	89408	89344	-	0
	Intr	89583	89508	-	0
	Intr	89751	89686	-	0
30	Intr	89916	89851	-	0
	Intr	90281	90192	-	0
	Init	90727	90584	-	0

>2342673 /39667

```

35      len =      827      nex =      2
      Init  95406  95717      +      0
      Term  95822  96232      +      0

```

40 >2342717 /13928

	len =	4710	nex =	16	
	Term	28916	28495	-	0
45	Intr	29102	29002	-	0
	Intr	29276	29211	-	0
	Intr	29479	29365	-	0
	Intr	29760	29654	-	0
	Intr	29937	29848	-	0
50	Intr	30204	30094	-	0
	Intr	30570	30505	-	0
	Intr	30730	30665	-	0
	Intr	31414	31265	-	0
	Intr	31587	31513	-	0
55	Intr	32170	32079	-	0
	Intr	32332	32267	-	0
	Intr	32516	32417	-	0
	Intr	32772	32611	-	0
60	Intr	33012	32912	-	0

>2342717 /23892

```
len = 1550    nex = 4
```

5	Term	33902	33442	-	0
	Intr	34398	34340	-	0
	Intr	34564	34485	-	0
	Init	34991	34651	-	0

10 >2342717 /25519

```
len = 2805    nex = 5
```

15	Term	38674	38181	-	0
	Intr	38927	38769	-	0
	Intr	39218	39037	-	0
	Intr	40474	40303	-	0
	Init	40985	40560	-	0

20 >2351061 /36048

```
len = 2257    nex = 4
```

	Term	36654	36150	-	0
25	Intr	37353	37320	-	0
	Intr	37883	37644	-	0
	Init	38406	38255	-	0

>2351061 /16286

```

30      len =      1302      nex =          2
      Init  60023  60178      +          0
      Term  60434  60780      +          0

```

>2351061 /25119

```
len = 2152    nex = 5
```

40	Inlt	72312	72460	+	0
	Intr	72978	73443	+	0
	Intr	73577	73670	+	0
	Intr	73763	73893	+	0
	Term	74106	74463	+	0

>2351061 /7022

```
len = 1348    nex = 1
```

50	Snql	74769	74513	-	0
----	------	-------	-------	---	---

>2351061 /37512

```
len = 1737    nex = 0
```

```

55      >2351062      /1575
      len =      1492      nex =      2
60      Init 11143 11366      +      0

```

952

	Term	11952	12270	+	0
	>2351062 /38092				
5	len =	2470	nex =	3	
	Term	27085	26904	-	0
	Intr	28828	27521	-	0
	Init	29365	29247	-	0
10	>2351062 /17241				
	len =	1404	nex =	3	
15	Init	29965	30040	+	0
	Intr	30233	30463	+	0
	Term	30712	30955	+	0
	>2351062 /31041				
20	len =	2710	nex =	8	
	Init	50901	51179	+	0
	Intr	51563	51664	+	0
25	Intr	51779	51832	+	0
	Intr	52010	52102	+	0
	Intr	52264	52356	+	0
	Intr	52687	52791	+	0
	Intr	52881	52979	+	0
30	Term	53072	53603	+	0
	>2351062 /23924				
	len =	1277	nex =	3	
35	Init	71481	71998	+	0
	Intr	72070	72397	+	0
	Term	72483	72757	+	0
40	>2351063 /114691				
	len =	1789	nex =	8	
	Term	20785	20575	-	0
45	Intr	20954	20889	-	0
	Intr	21132	21047	-	0
	Intr	21269	21235	-	0
	Intr	21455	21369	-	0
	Intr	21616	21539	-	0
50	Intr	21741	21701	-	0
	Init	22363	22239	-	0
	>2351063 /36626				
55	len =	1476	nex =	6	
	Term	21132	21053	-	0
	Intr	21269	21235	-	0
	Intr	21455	21369	-	0
60	Intr	21616	21539	-	0

2025 RELEASE UNDER E.O. 14176

953

	Intr	21741	21701	-	0
	Init	22528	22239	-	0
5 >2351063 /31913					
	len =	1211	nex =	4	
	Init	28196	28319	+	0
10	Intr	28394	28464	+	0
	Intr	28552	28573	+	0
	Term	28658	29015	+	0
15 >2351063 /103246					
	len =	1195	nex =	4	
	Init	28196	28319	+	0
	Intr	28394	28464	+	0
20	Intr	28552	28573	+	0
	Term	28658	29015	+	0
25 >2351063 /36058					
	len =	2835	nex =	10	
	Init	55242	55559	+	0
	Intr	55634	55699	+	0
	Intr	55825	55890	+	0
30	Intr	56186	56264	+	0
	Intr	56488	56608	+	0
	Intr	56694	56789	+	0
	Intr	56864	56976	+	0
	Intr	57238	57354	+	0
35	Intr	57635	57735	+	0
	Term	57871	58076	+	0
40 >2351063 /95281					
	len =	430	nex =	1	
	Sngl	58996	59416	+	0
45 >2351063 /108981					
	len =	314	nex =	1	
	Sngl	62819	63132	+	0
50 >2351063 /19716					
	len =	2088	nex =	8	
	Term	66456	66175	-	0
	Intr	66816	66527	-	0
55	Intr	67192	66895	-	0
	Intr	67350	67280	-	0
	Intr	67560	67444	-	0
	Intr	67709	67635	-	0
	Intr	67857	67796	-	0
60	Init	68262	68028	-	0

>2351063 /18140

```

5      len =      3071      nex =      5
      Term 81242  80797      -      0
      Intr 81474  81378      -      0
      Intr 81610  81555      -      0
      Intr 81979  81686      -      0
10     Init 82808  82071      -      0

```

>2351064 /10154

15	len =	1092	nex =	5	
	Init	30526	30610	+	0
	Intr	30871	30941	+	0
	Intr	31032	31188	+	0
	Intr	31364	31450	+	0
20	Term	31536	31617	+	0

>2351064 /23922

25	len =	2156	nex =	9	
	Init	30531	30610	+	0
	Intr	30871	30941	+	0
	Intr	31032	31188	+	0
	Intr	31364	31450	+	0
30	Intr	31536	31687	+	0
	Intr	31802	31882	+	0
	Intr	31983	32091	+	0
	Intr	32233	32359	+	0
	Term	32454	32686	+	0

>2351064 /41054

```

len =      500    nex =      2
40      Init  32229  32359      +      0
      Term  32454  32728      +      0

```

>2351064 /37122

45	len =	2271	nex =	5	
	Term	52016	51678	-	0
	Intr	52304	52104	-	0
	Intr	52616	52417	-	0
50	Intr	52811	52698	-	0
	Init	53187	53050	-	0

>2351065 /8508

```
55      len =      286      nex =      1
          Snq1 1156      871      -      0
```

>2351065 /29363

60

955

	len =	8125	nex =	3	
	Term	5274	4953	-	0
	Intr	12650	5804	-	0
5	Init	13070	12743	-	0
	>2351065	/3542			
	len =	1606	nex =	3	
10	Term	12650	12382	-	0
	Intr	13557	12743	-	0
	Init	13987	13679	-	0
15	>2351065	/117588			
	len =	1433	nex =	3	
	Init	26825	26985	+	0
20	Intr	27076	27149	+	0
	Term	27414	28257	+	0
	>2351065	/15229			
25	len =	1952	nex =	9	
	Term	28953	28676	-	0
	Intr	29086	29035	-	0
30	Intr	29404	29169	-	0
	Intr	29662	29605	-	0
	Intr	29821	29753	-	0
	Intr	30022	29914	-	0
	Intr	30232	30165	-	0
	Intr	30434	30315	-	0
35	Init	30627	30561	-	0
	>2351065	/41047			
	len =	1956	nex =	9	
40	Term	28953	28675	-	0
	Intr	29086	29035	-	0
	Intr	29404	29169	-	0
	Intr	29662	29605	-	0
45	Intr	29821	29753	-	0
	Intr	30022	29914	-	0
	Intr	30232	30165	-	0
	Intr	30434	30315	-	0
	Init	30630	30561	-	0
50	>2351065	/105944			
	len =	254	nex =	1	
55	Sngl	38996	38743	-	0
	>2351065	/6823			
60	len =	436	nex =	1	

956

	Sngl	420	855	+	0
	>2351065 /15640				
5	len =	2139	nex =	7	
	Term	54303	53997	-	0
	Intr	54528	54415	-	0
	Intr	54773	54648	-	0
10	Intr	55027	54948	-	0
	Intr	55198	55117	-	0
	Intr	55390	55316	-	0
	Init	56135	55791	-	0
15	>2351065 /633				
	len =	529	nex =	1	
	Sngl	56522	57050	+	0
20	>2351065 /104017				
	len =	1017	nex =	2	
25	Term	62259	61832	-	0
	Init	62503	62277	-	0
	>2351066 /92216				
30	len =	1063	nex =	1	
	Sngl	2252	1951	-	0
	>2351066 /18332				
35	len =	372	nex =	1	
	Sngl	51067	50696	-	0
40	>2351066 /19255				
	len =	1001	nex =	2	
	Init	6275	6505	+	0
45	Term	6677	6809	+	0
	>2351066 /93148				
	len =	557	nex =	1	
50	Sngl	64963	64407	-	0
	>2351066 /9184				
55	len =	1493	nex =	3	
	Init	65437	65484	+	0
	Intr	65563	65622	+	0
	Term	66328	66800	+	0
60					

2025 RELEASE UNDER E.O. 14176

957

>2351066 /94924

len = 772 nex = 4

5	Term	66989	66757	-	0
	Intr	67176	67069	-	0
	Intr	67314	67274	-	0
	Init	67528	67424	-	0

10 >2351066 /117503

len = 1270 nex = 5

15	Term	82968	82813	-	0
	Intr	83338	83123	-	0
	Intr	83553	83453	-	0
	Intr	83928	83699	-	0
	Init	84064	83998	-	0

20 >2351067 /24137

len = 592 nex = 1

25	Sngl	23773	23182	-	0
----	------	-------	-------	---	---

>2351067 /102435

len = 1553 nex = 1

30	Sngl	31407	31589	+	0
----	------	-------	-------	---	---

>2351067 /42506

len = 913 nex = 2

35	Init	3624	3761	+	0
	Term	4109	4536	+	0

>2351067 /37503

len = 1398 nex = 4

40	Term	39519	39286	-	0
	Intr	39736	39638	-	0
45	Intr	40371	40283	-	0
	Init	40683	40599	-	0

>2351067 /23800

len = 1450 nex = 4

50	Term	39519	39294	-	0
	Intr	39736	39638	-	0
	Intr	40371	40283	-	0
55	Init	40735	40599	-	0

>2351067 /12458

len = 192 nex = 1

60

958

	Sngl	43705	43896	+	0
	>2351068 /108814				
5	len =	755	nex =	4	
	Init	14299	14392	+	0
	Intr	14508	14644	+	0
	Intr	14771	14817	+	0
10	Term	14906	15053	+	0
	>2351068 /33315				
	len =	2311	nex =	9	
15	Init	14309	14392	+	0
	Intr	14508	14644	+	0
	Intr	14771	14817	+	0
	Intr	14906	15231	+	0
20	Intr	15511	15593	+	0
	Intr	15693	15768	+	0
	Intr	15855	16012	+	0
	Intr	16102	16263	+	0
25	Term	16357	16619	+	0
	>2351068 /37265				
	len =	2272	nex =	9	
30	Init	14347	14392	+	0
	Intr	14508	14644	+	0
	Intr	14771	14817	+	0
	Intr	14906	15231	+	0
	Intr	15511	15593	+	0
35	Intr	15693	15768	+	0
	Intr	15855	16012	+	0
	Intr	16102	16263	+	0
	Term	16357	16618	+	0
40	>2351068 /777				
	len =	540	nex =	1	
	Sngl	22901	23440	+	0
45	>2351068 /2304				
	len =	550	nex =	1	
50	Sngl	22901	23442	+	0
	>2351068 /15211				
	len =	560	nex =	1	
55	Sngl	22904	23463	+	0
	>2351068 /27372				
60	len =	1870	nex =	3	

ORIGINAL PAGE NO. 00000000

	Init	42505	42957	+	0
	Intr	43205	43414	+	0
	Term	43963	44371	+	0
5	>2351068 /5335				
	len =	731	nex =	3	
10	Init	5218	5304	+	0
	Intr	5320	5477	+	0
	Term	5551	5931	+	0
	>2351068 /22794				
15	len =	857	nex =	1	
	Sngl	61140	61996	+	0
20	>2351068 /28601				
	len =	2200	nex =	7	
	Init	65723	65950	+	0
25	Intr	66035	66198	+	0
	Intr	66298	66349	+	0
	Intr	66544	66771	+	0
	Intr	66874	67063	+	0
	Intr	67153	67418	+	0
30	Term	67680	67922	+	0
	>2351068 /25211				
	len =	1220	nex =	1	
35	Sngl	69965	68746	-	0
	>2351069 /26016				
40	len =	2002	nex =	7	
	Init	26231	26670	+	0
	Intr	26762	26870	+	0
	Intr	26960	27122	+	0
45	Intr	27209	27357	+	0
	Intr	27450	27601	+	0
	Intr	27686	27800	+	0
	Term	27886	28232	+	0
50	>2351069 /1271				
	len =	3480	nex =	9	
	Init	42775	42864	+	0
55	Intr	43235	43369	+	0
	Intr	43517	43633	+	0
	Intr	43791	43942	+	0
	Intr	44014	44098	+	0
	Intr	44277	44371	+	0
60	Intr	44852	45017	+	0

960

Intr	45150	45345	+	0
Term	45434	45819	+	0

Term	45434	45819	+	0
------	-------	-------	---	---

>2351069 /13271

```

5      len =      702    nex =      3

```

Init	62856	62885	+	0
------	-------	-------	---	---

Intr	62964	63042	+	0
------	-------	-------	---	---

	Incl	62584	63042		
10	Term	63127	63557	+	0

>2351069 /7744

```
len = 1618    nex = 8
```

Term	67305	66948	-	0
------	-------	-------	---	---

Intr	67508	67411	-	0
------	-------	-------	---	---

Intr	67723	67598	-	0
------	-------	-------	---	---

Intr	67896	67813	-	0
------	-------	-------	---	---

20	Intr	68098	67982	-	0
----	------	-------	-------	---	---

Intr	68261	68178	-	0
------	-------	-------	---	---

Intr	68427	68380	-	0
------	-------	-------	---	---

Init	68562	68508	-	0
------	-------	-------	---	---

25 >2351069 /3285

```
len = 3163    nex = 13
```

Term	67262	67061	-	0
------	-------	-------	---	---

30	Intr	67508	67411	-	0
----	------	-------	-------	---	---

Intr	67723	67598	-	0
------	-------	-------	---	---

Intr	67896	67813	-	0
------	-------	-------	---	---

Intr	68098	67982	-	0
------	-------	-------	---	---

Intr	68261	68178	-	0
------	-------	-------	---	---

35	Intr	68427	68380	-	0
----	------	-------	-------	---	---

Intr	68562	68508	-	0
------	-------	-------	---	---

Intr	68759	68704	-	0
------	-------	-------	---	---

Intr	68928	68844	-	0
------	-------	-------	---	---

Intr	69102	69029	-	0
------	-------	-------	---	---

40	Intr	69415	69349	-	0
----	------	-------	-------	---	---

Init	70098	70008	-	0
------	-------	-------	---	---

>2351070 /97197

```
45      len =      697      nex =      1
```

Snq1	23957	23261	-	0
------	-------	-------	---	---

>2351070 /6363

```
50      len =      560    nex =      1
```

Snq1	34956	34397	-	0
------	-------	-------	---	---

55 >2351070 /26053

```
len =      817      nex =      1
```

Snq1	46123	46936	+	0
------	-------	-------	---	---

60

>2351071 /17432

len = 2313 nex = 9

5	Term	46885	46586	-	0
	Intr	47174	47088	-	0
	Intr	47356	47291	-	0
	Intr	47556	47467	-	0
	Intr	47720	47640	-	0
10	Intr	47910	47833	-	0
	Intr	48093	48003	-	0
	Intr	48436	48295	-	0
	Init	48898	48628	-	0

15 >2351071 /39195

len = 2186 nex = 3

20	Term	70730	70227	-	0
	Intr	71606	71158	-	0
	Init	72412	72145	-	0

>2351071 /17360

25 len = 1402 nex = 3

	Term	78193	77927	-	0
	Intr	78535	78274	-	0
	Init	79311	79168	-	0

30 >2351071 /26743

len = 1466 nex = 3

35	Term	78193	77927	-	0
	Intr	78535	78274	-	0
	Init	79392	79168	-	0

>2351072 /29659

40 len = 2508 nex = 5

	Term	22869	22279	-	0
	Intr	23128	23019	-	0
45	Intr	23667	23238	-	0
	Intr	23978	23838	-	0
	Init	24786	24671	-	0

>2351072 /207148

50 len = 797 nex = 1

Sngl 50991 50195 - 0

55 >2351073 /98326

len = 676 nex = 3

60	Term	19588	19334	-	0
	Intr	19757	19681	-	0

05030300-061600

962

	Init	19996	19838	-	0
	>2351073 /100141				
5	len =	1717	nex =	5	
	Term	19588	19293	-	0
	Intr	19757	19681	-	0
	Intr	20220	19838	-	0
10	Intr	20633	20533	-	0
	Init	21009	20902	-	0
	>2351073 /115914				
15	len =	116	nex =	1	
	Sngl	26710	26595	-	0
	>2351073 /95599				
20	len =	749	nex =	3	
	Term	26967	26608	-	0
	Intr	27178	27047	-	0
25	Init	27356	27258	-	0
	>2351073 /35552				
	len =	1828	nex =	6	
30	Term	26967	26653	-	0
	Intr	27178	27047	-	0
	Intr	27399	27258	-	0
	Intr	27742	27550	-	0
35	Intr	28087	27842	-	0
	Init	28480	28170	-	0
	>2351073 /118777				
40	len =	1030	nex =	1	
	Sngl	31871	32900	+	0
	>2358139 /20380				
45	len =	876	nex =	3	
	Init	15794	15936	+	0
	Intr	16035	16176	+	0
50	Term	16428	16669	+	0
	>2358139 /29808				
	len =	1270	nex =	2	
55	Term	64249	63873	-	0
	Init	65100	64760	-	0
	>2358139 /108558				
60					

963

	len =	1069	nex =	3	
	Init	65271	65413	+	0
	Intr	65781	65860	+	0
5	Term	66116	66339	+	0
	>2358139 /1730				
	len =	1484	nex =	3	
10	Init	71725	71848	+	0
	Intr	72291	72590	+	0
	Term	72701	73208	+	0
15	>2392762 /8805				
	len =	1259	nex =	2	
	Term	30586	29909	-	0
20	Init	31167	30868	-	0
	>2392762 /14724				
	len =	1796	nex =	8	
25	Term	60877	60621	-	0
	Intr	61051	60973	-	0
	Intr	61293	61140	-	0
	Intr	61514	61420	-	0
30	Intr	61620	61585	-	0
	Intr	61952	61727	-	0
	Intr	62107	62037	-	0
	Init	62416	62342	-	0
35	>2392762 /15990				
	len =	1729	nex =	8	
	Term	60877	60688	-	0
40	Intr	61051	60973	-	0
	Intr	61293	61140	-	0
	Intr	61514	61420	-	0
	Intr	61620	61585	-	0
	Intr	61952	61727	-	0
45	Intr	62107	62037	-	0
	Init	62416	62342	-	0
	>2392762 /41162				
50	len =	951	nex =	3	
	Init	68249	68350	+	0
	Intr	68449	68513	+	0
	Term	68901	69199	+	0
55	>2435510 /32833				
	len =	1450	nex =	5	
60	Term	41015	40654	-	0

2025 RELEASE UNDER E.O. 14176

964

	Intr	41265	41098	-	0
	Intr	41451	41368	-	0
	Intr	41718	41540	-	0
	Init	42097	41892	-	0
5	>2435510 /1011				
	len =	1120	nex =	3	
10	Term	51801	51490	-	0
	Intr	52028	51949	-	0
	Init	52609	52122	-	0
15	>2435510 /19362				
	len =	1041	nex =	3	
	Init	61031	61254	+	0
	Intr	61359	61535	+	0
20	Term	61610	62071	+	0
25	>2435510 /142314				
	len =	919	nex =	3	
	Init	61151	61254	+	0
	Intr	61359	61535	+	0
	Term	61610	62069	+	0
30	>2435510 /33456				
	len =	2142	nex =	6	
	Term	4364	4051	-	0
35	Intr	4676	4612	-	0
	Intr	5214	5151	-	0
	Intr	5423	5314	-	0
	Intr	5600	5513	-	0
	Init	6192	5794	-	0
40	>2435510 /4367				
	len =	2039	nex =	8	
45	Init	76018	76119	+	0
	Intr	76377	76574	+	0
	Intr	76648	76707	+	0
	Intr	76793	77235	+	0
	Intr	77335	77501	+	0
50	Intr	77587	77660	+	0
	Intr	77749	77808	+	0
	Term	77912	78053	+	0
55	>2443899 /22008				
	len =	1489	nex =	2	
	Term	102074	101797	-	0
	Init	103282	102296	-	0
60					

2025 RELEASE UNDER E.O. 14176

>2443899 /1734

len = 888 nex = 2

5	Term	14747	14318	-	0
	Init	15205	14965	-	0

>2459406 /42992

10 len = 2396 nex = 10

	Term	117911	117825	-	0
	Intr	118071	117986	-	0
	Intr	118340	118166	-	0
15	Intr	118518	118458	-	0
	Intr	118661	118595	-	0
	Intr	118838	118754	-	0
	Intr	119077	118920	-	0
	Intr	119310	119166	-	0
20	Intr	119486	119427	-	0
	Init	119855	119575	-	0

>2459406 /11254

25 len = 2035 nex = 6

	Init	128392	128598	+	0
	Intr	128894	129063	+	0
	Intr	129142	129327	+	0
30	Intr	129412	129577	+	0
	Intr	129681	129870	+	0
	Term	130089	130426	+	0

>2459406 /92741

35 len = 538 nex = 1

Sngl 141230 140693 - 0

40 >2459406 /13741

len = 1713 nex = 4

	Term	18475	18146	-	0
45	Intr	18628	18567	-	0
	Intr	19123	18713	-	0
	Init	19858	19394	-	0

>2459406 /25272

50 len = 1750 nex = 4

	Init	2679	2985	+	0
	Intr	3377	3419	+	0
55	Intr	3511	3571	+	0
	Term	3697	4419	+	0

>2459406 /35273

60 len = 2218 nex = 3

	Term	26889	26777	-	0
	Intr	28208	27837	-	0
	Init	28994	28459	-	0
5	>2459406 /28563				
	len =	1150	nex =	4	
10	Term	47656	47428	-	0
	Intr	47792	47751	-	0
	Intr	48158	47874	-	0
	Init	48577	48488	-	0
15	>2459406 /119409				
	len =	468	nex =	1	
20	Sngl 57470 57023 - 0				
	>2459406 /116034				
	len =	337	nex =	1	
25	Sngl 61222 61558 + 0				
	>2459406 /8717				
	len =	2113	nex =	4	
30	Init	66546	66940	+	0
	Intr	67084	67181	+	0
	Intr	67274	67339	+	0
	Term	68443	68658	+	0
35	>2459406 /31633				
	len =	945	nex =	2	
40	Init	77435	77674	+	0
	Term	78004	78379	+	0
	>2459406 /19302				
45	len =	2115	nex =	6	
	Term	80490	80306	-	0
	Intr	80717	80586	-	0
	Intr	80949	80814	-	0
50	Intr	81174	81044	-	0
	Intr	81479	81424	-	0
	Init	82420	82270	-	0
	>2459406 /37919				
55	len =	2274	nex =	6	
	Term	80490	80262	-	0
	Intr	80717	80586	-	0
60	Intr	80949	80814	-	0

967

	Intr	81174	81044	-	0
	Intr	81479	81424	-	0
	Init	82535	82270	-	0
5	>2459406 /18894				
	len =	235	nex =	1	
10	Sngl	85070	85304	+	0
	>2477521 /15308				
	len =	1434	nex =	1	
15	Sngl	11192	12625	+	0
	>2477521 /27205				
	len =	760	nex =	3	
20	Term	22663	22447	-	0
	Intr	22864	22743	-	0
	Init	23206	22955	-	0
25	>2477521 /40049				
	len =	3210	nex =	5	
30	Init	52491	52536	+	0
	Intr	52618	52732	+	0
	Intr	52824	52891	+	0
	Intr	52986	53708	+	0
	Term	53792	54336	+	0
35	>2477521 /3549				
	len =	1750	nex =	4	
40	Init	59783	60056	+	0
	Intr	60329	60677	+	0
	Intr	60773	60914	+	0
	Term	60979	61527	+	0
	>2477521 /12293				
45	len =	1796	nex =	6	
	Term	71123	70636	-	0
	Intr	71380	71205	-	0
50	Intr	71502	71478	-	0
	Intr	71702	71620	-	0
	Intr	72024	71951	-	0
	Init	72431	72108	-	0
55	>2477521 /98850				
	len =	4463	nex =	7	
60	Init	74583	74814	+	0
	Intr	77407	77441	+	0

968

	Intr	77553	77614	+	0
	Intr	77696	77795	+	0
	Intr	77904	77945	+	0
	Intr	78281	78322	+	0
5	Term	78695	79045	+	0
>2477521 /92459					
10	len =	4460	nex =	7	
	Init	74588	74814	+	0
	Intr	77285	77342	+	0
	Intr	77553	77614	+	0
	Intr	77696	77795	+	0
15	Intr	77904	77945	+	0
	Intr	78281	78322	+	0
	Term	78695	79047	+	0
>2477521 /5076					
20	len =	730	nex =	3	
	Term	79591	79372	-	0
	Intr	79924	79697	-	0
25	Init	80096	80042	-	0
>2477521 /4033					
30	len =	1930	nex =	7	
	Init	94403	94493	+	0
	Intr	94625	94761	+	0
	Intr	94865	94911	+	0
	Intr	94999	95483	+	0
35	Intr	95570	95727	+	0
	Intr	95814	95975	+	0
	Term	96051	96327	+	0
>2494106 /36412					
40	len =	1375	nex =	3	
	Term	99606	98923	-	0
	Intr	100124	99692	-	0
45	Init	100297	100214	-	0
>2494106 /11408					
50	len =	644	nex =	1	
	Sngl	109531	110174	+	0
>2494106 /8951					
55	len =	910	nex =	1	
	Sngl	112974	112773	-	0
>2494106 /37020					
60					

20250927 14:23:50

969

	len =	757	nex =	4	
	Term	122980	122712	-	0
	Intr	123133	123078	-	0
5	Intr	123278	123220	-	0
	Init	123468	123370	-	0
>2494106 /29872					
10	len =	861	nex =	2	
	Term	122980	122712	-	0
	Init	123133	123078	-	0
15	>2494106 /34434				
	len =	866	nex =	4	
	Term	122980	122714	-	0
20	Intr	123133	123078	-	0
	Intr	123278	123220	-	0
	Init	123577	123370	-	0
>2494106 /34374					
25	len =	359	nex =	2	
	Term	123278	123219	-	0
	Init	123577	123370	-	0
30	>2494106 /5465				
	len =	2050	nex =	7	
35	Init	132597	132734	+	0
	Intr	133129	133207	+	0
	Intr	133336	133389	+	0
	Intr	133680	133793	+	0
	Intr	134040	134107	+	0
40	Intr	134190	134301	+	0
	Term	134381	134640	+	0
>2494106 /1520					
45	len =	1810	nex =	6	
	Init	132677	133207	+	0
	Intr	133336	133389	+	0
	Intr	133680	133793	+	0
50	Intr	134040	134107	+	0
	Intr	134190	134301	+	0
	Term	134381	134477	+	0
>2494106 /2681					
55	len =	910	nex =	1	
	Sngl	143514	143911	+	0
60	>2494106 /33770				

2025 RELEASE UNDER E.O. 14176

	len =	1302	nex =	4	
	Term	158712	158351	-	0
5	Intr	159059	158976	-	0
	Intr	159236	159156	-	0
	Init	159509	159332	-	0
	>2494106 /27457				
10	len =	1462	nex =	4	
	Term	40898	40547	-	0
	Intr	41137	41003	-	0
15	Intr	41443	41231	-	0
	Init	42008	41526	-	0
	>2494106 /25255				
20	len =	1719	nex =	3	
	Init	54004	54063	+	0
	Intr	54151	54486	+	0
	Term	54639	54877	+	0
25	>2494106 /14939				
	len =	610	nex =	2	
30	Init	54277	54486	+	0
	Term	54639	54879	+	0
	>2494106 /32130				
35	len =	3130	nex =	12	
	Term	56042	55686	-	0
	Intr	56181	56114	-	0
	Intr	56328	56265	-	0
40	Intr	56502	56421	-	0
	Intr	56676	56618	-	0
	Intr	56984	56925	-	0
	Intr	57266	57104	-	0
	Intr	57498	57374	-	0
45	Intr	57857	57795	-	0
	Intr	58060	58001	-	0
	Intr	58325	58140	-	0
	Init	58811	58689	-	0
50	>2494106 /6667				
	len =	1554	nex =	1	
	Sngl	60644	59091	-	0
55	>2494106 /25894				
	len =	1630	nex =	4	
60	Term	64139	63599	-	0

971

Intr	64439	64381	-	0
Intr	64965	64855	-	0
Init	65226	65138	-	0

5 >2494110 /23300

len = 2036 nex = 5

Term	17775	17469	-	0
Intr	18041	17877	-	0
Intr	18302	18159	-	0
Intr	18618	18423	-	0
Init	19504	19053	-	0

15 >2494110 /8559

len = 1302 nex = 2

Init	25200	25402	+	0
Term	26210	26501	+	0

>2494110 /37952

25 len = 4214 nex = 6

Init	25200	25402	+	0
Intr	26210	26290	+	0
Intr	27617	28259	+	0
Intr	28358	28461	+	0
Intr	28571	28709	+	0
Term	28803	29413	+	0

>2494110 /21100

35 len = 812 nex = 3

Term	30699	30410	-	0
Intr	30921	30796	-	0
Init	31221	30993	-	0

40 >2494110 /34753

len = 807 nex = 3

Term	30699	30415	-	0
Intr	30921	30796	-	0
Init	31221	30993	-	0

>2494110 /110726

50 len = 493 nex = 1

Sngl 32194 32672 + 0

55 >2494110 /2265

len = 494 nex = 1

Sngl 38819 39312 + 0

60

972

>2494110 /13232

```
len = 1220    nex = 1
```

5	Sngl	40544	39752	-	0
---	------	-------	-------	---	---

>2494110 /31923

```
len = 1284    nex = 3
```

10	Init	41985	42310	+	0
	Intr	42859	42930	+	0
	Term	43017	43268	+	0

15 >2494110 /100984

```
len =      1340No match - No prediction
>2494110      /27110
```

```

20      len =      108      nex =      1
      Sngl 74373 74480      +      0

```

>2494110 /40608

```
len = 1703    nex = 5
```

	Term	91321	90966	-	0
	Intr	91466	91405	-	0
30	Intr	91657	91540	-	0
	Intr	92025	91739	-	0
	Init	92668	92298	-	0

>2494110 /2935

33	len =	1613	nex =	5	
	Init	97175	97627	+	0
	Intr	97725	97897	+	0
40	Intr	97974	98088	+	0
	Intr	98324	98478	+	0
	Term	98578	98787	+	0

>2505864 /35333

```
len = 1426    nex = 4
```

	Init	20951	21020	+	0
	Intr	21255	21415	+	0
50	Intr	21681	21869	+	0
	Term	22136	22367	+	0

>2505864 /4328

```
55      len =      1374      nex =          4
```

	Init	21013	21070	+	0
	Intr	21255	21415	+	0
	Intr	21681	21869	+	0
60	Term	22136	22386	+	0

>2505873 /6115

5	len =	1600	nex =	2	
	Term	14088	13696	-	0
	Init	15295	14419	-	0

>2505873 /33852

10	len =	134	nex =	1	
	Sngl	19922	20055	+	0

15 >2505873 /36699

	len =	236	nex =	1	
	Sngl	27483	27718	+	0

20 >2529657 /32457

	len =	1232	nex =	5	
25	Term	10325	10206	-	0
	Intr	10512	10408	-	0
	Intr	10777	10703	-	0
	Intr	11135	11111	-	0
	Init	11437	11243	-	0

30 >2529657 /26123

	len =	1422	nex =	5	
35	Term	10325	10041	-	0
	Intr	10512	10408	-	0
	Intr	10777	10703	-	0
	Intr	11135	11111	-	0
	Init	11462	11380	-	0

40 >2529657 /20647

	len =	1390	nex =	5	
45	Term	12109	11630	-	0
	Intr	12283	12185	-	0
	Intr	12499	12362	-	0
	Intr	12722	12592	-	0
	Init	13015	12840	-	0

50 >2529657 /28691

	len =	2057	nex =	7	
55	Term	12109	11913	-	0
	Intr	12283	12185	-	0
	Intr	12499	12362	-	0
	Intr	12722	12592	-	0
	Intr	12986	12840	-	0
60	Intr	13647	13615	-	0

005522-001600

	Init	13969	13735	-	0
	>2529657 /33373				
5	len =	2492	nex =	8	
	Term	12109	11637	-	0
	Intr	12283	12185	-	0
	Intr	12499	12362	-	0
10	Intr	12722	12592	-	0
	Intr	12986	12840	-	0
	Intr	13647	13615	-	0
	Intr	13831	13735	-	0
	Init	14128	13974	-	0
15	>2529657 /24272				
	len =	1054	nex =	4	
20	Term	17370	17243	-	0
	Intr	17555	17463	-	0
	Intr	17935	17637	-	0
	Init	18296	18094	-	0
25	>2529657 /6394				
	len =	1870	nex =	5	
	Term	17370	16988	-	0
30	Intr	17555	17463	-	0
	Intr	17935	17637	-	0
	Intr	18295	18094	-	0
	Init	18459	18415	-	0
35	>2529657 /25729				
	len =	802	nex =	1	
40	Sngl	3834	4635	+	0
	>2529657 /37870				
	len =	3805	nex =	17	
45	Term	46706	46424	-	0
	Intr	46947	46882	-	0
	Intr	47087	47058	-	0
	Intr	47280	47182	-	0
	Intr	47466	47371	-	0
50	Intr	47623	47573	-	0
	Intr	47773	47707	-	0
	Intr	47950	47856	-	0
	Intr	48158	48077	-	0
	Intr	48324	48275	-	0
55	Intr	48463	48413	-	0
	Intr	48638	48540	-	0
	Intr	49052	48969	-	0
	Intr	49302	49192	-	0
	Intr	49575	49426	-	0
60	Intr	49795	49678	-	0

2025 RELEASE UNDER E.O. 14176

975

	Init	50050	49884	-	0
	>2529657	/32039			
5	len =	670	nex =	1	
	Sngl	63987	64654	+	0
	>2529657	/9499			
10	len =	654	nex =	2	
	Term	65131	64658	-	0
	Init	65297	65222	-	0
15	>2529657	/38461			
	len =	2350	nex =	10	
20	Term	65131	64823	-	0
	Intr	65346	65222	-	0
	Intr	65588	65432	-	0
	Intr	65777	65686	-	0
	Intr	65890	65863	-	0
25	Intr	66093	65976	-	0
	Intr	66394	66339	-	0
	Intr	66604	66507	-	0
	Intr	66777	66693	-	0
	Init	67165	66986	-	0
30	>2529657	/13774			
	len =	2397	nex =	10	
35	Term	65131	64823	-	0
	Intr	65346	65222	-	0
	Intr	65588	65432	-	0
	Intr	65777	65686	-	0
	Intr	65890	65863	-	0
40	Intr	66093	65976	-	0
	Intr	66394	66339	-	0
	Intr	66604	66507	-	0
	Intr	66777	66693	-	0
	Init	67219	66986	-	0
45	>2529657	/34914			
	len =	717	nex =	1	
50	Sngl	75255	74539	-	0
	>2529657	/37980			
	len =	1352	nex =	1	
55	Sngl	75893	74542	-	0
	>2564044	/156017			
60	len =	401	nex =	1	

2025 RELEASE UNDER E.O. 14176

976

	Sngl	12975	12575	-	0
5	>2564044	/156773			
	len =	350	nex =	1	
	Sngl	12997	12648	-	0
10	>2564044	/31129			
	len =	430	nex =	1	
	Sngl	13041	12616	-	0
15	>2564044	/21629			
	len =	1610	nex =	5	
20	Term	36986	36739	-	0
	Intr	37123	37068	-	0
	Intr	37318	37272	-	0
	Intr	37669	37626	-	0
	Init	38348	38232	-	0
25	>2564044	/22860			
	len =	3400	nex =	11	
30	Init	5043	5315	+	0
	Intr	5670	5734	+	0
	Intr	5871	5969	+	0
	Intr	6171	6303	+	0
	Intr	6748	6807	+	0
35	Intr	6897	7019	+	0
	Intr	7379	7450	+	0
	Intr	7562	7699	+	0
	Intr	7786	7941	+	0
	Intr	8028	8132	+	0
40	Term	8282	8442	+	0
	>2564045	/108335			
	len =	1516	nex =	2	
45	Term	653	118	-	0
	Init	1633	770	-	0
	>2564045	/512			
50	len =	1435	nex =	1	
	Sngl	40196	39668	-	0
55	>2564045	/40250			
	len =	1210	nex =	2	
60	Term	57008	56234	-	0
	Init	57441	57096	-	0

977

>2564045 /36090

```

5      len =      1219      nex =          2
      Term  57008    56234      -          0
      Init  57452    57096      -          0

```

>2564045 /33763

```

10      len =      1217      nex =          1
      Sngl  5886      7102      +          0

```

15 >2564045 /23566

	len =	1043	nex =	3	
20	Init	9042	9192	+	0
	Intr	9618	9763	+	0
	Term	9851	10084	+	0

>2564046 /4272

25	len =	4185	nex =	11	
	Term	18249	17894	-	0
	Intr	18506	18454	-	0
30	Intr	18683	18598	-	0
	Intr	18985	18867	-	0
	Intr	19502	19431	-	0
	Intr	19881	19708	-	0
	Intr	20444	20289	-	0
	Intr	20917	20836	-	0
35	Intr	21276	21130	-	0
	Intr	21654	21468	-	0
	Init	22078	21842	-	0

>2564046 /13993

40	len =	1672	nex =	6	
	Init	27089	27339	+	0
	Intr	27573	27725	+	0
45	Intr	27820	27972	+	0
	Intr	28179	28262	+	0
	Intr	28344	28485	+	0
	Term	28581	28760	+	0

50 >2564046 /35683

```

len =      697  nex =      3
55  Term 34417 34208 - 0
    Intr 34609 34504 - 0
    Init 34904 34742 - 0

```

>2564047 /12802

```
60      len =      1648      nex =          3
```

978

	Init	16402	16518	+	0
	Intr	17081	17129	+	0
	Term	17663	17714	+	0
5	>2564047	/19442			
	len =	850	nex =	1	
10	Sngl	21861	22701	+	0
	>2564047	/2533			
	len =	1779	nex =	3	
15	Init	37480	37886	+	0
	Intr	37970	38637	+	0
	Term	39199	39258	+	0
20	>2564047	/32890			
	len =	1279	nex =	1	
	Sngl	51389	50111	-	0
25	>2564047	/13737			
	len =	1302	nex =	5	
30	Term	57880	57668	-	0
	Intr	58070	58011	-	0
	Intr	58297	58197	-	0
	Intr	58633	58398	-	0
	Init	58969	58725	-	0
35	>2564047	/6893			
	len =	1309	nex =	5	
40	Term	57880	57662	-	0
	Intr	58070	58011	-	0
	Intr	58297	58197	-	0
	Intr	58633	58398	-	0
	Init	58970	58725	-	0
45	>2564047	/114864			
	len =	2470	nex =	9	
50	Init	59318	59464	+	0
	Intr	59652	59723	+	0
	Intr	59821	59895	+	0
	Intr	60508	60588	+	0
	Intr	60854	60923	+	0
55	Intr	60996	61087	+	0
	Intr	61178	61219	+	0
	Intr	61298	61378	+	0
	Term	61566	61785	+	0
60	>2564047	/105566			

2025 RELEASE UNDER E.O. 14176

	len =	979	nex =	1	
5	Sngl	62070	63048	+	0
	>2564047	/12455			
	len =	1933	nex =	2	
10	Term	67046	66415	-	0
	Init	68347	67916	-	0
	>2564047	/40711			
15	len =	850	nex =	1	
	Sngl	78369	77529	-	0
	>2564048	/105906			
20	len =	1212	nex =	2	
	Init	2380	2769	+	0
	Term	2946	3591	+	0
25	>2564048	/115613			
	len =	586	nex =	1	
30	Sngl	31514	30929	-	0
	>2564048	/1200			
	len =	1778	nex =	4	
35	Term	38609	37885	-	0
	Intr	38864	38681	-	0
	Intr	39244	38988	-	0
	Init	39662	39331	-	0
40	>2564048	/39462			
	len =	2351	nex =	7	
45	Init	41518	41825	+	0
	Intr	42059	42268	+	0
	Intr	42387	42557	+	0
	Intr	42766	42889	+	0
	Intr	43155	43216	+	0
50	Init	43305	43386	+	0
	Term	43481	43868	+	0
	>2564048	/10292			
55	len =	1951	nex =	7	
	Init	41667	41825	+	0
	Intr	42059	42268	+	0
	Intr	42387	42557	+	0
60	Intr	42766	42889	+	0

980

Intr	43155	43216	+	0
Intr	43305	43386	+	0
Term	43481	43617	+	0

5 >2564048 /26637

len = 1980 nex = 7

10	Init	61938	62027	+	0
	Intr	62306	62497	+	0
	Intr	62586	62757	+	0
	Intr	62859	62932	+	0
	Intr	63011	63037	+	0
15	Intr	63126	63149	+	0
	Term	63245	63657	+	0

>2564048 /158431

len = 1690 nex = 5

20	Init	65258	65519	+	0
	Intr	65699	65751	+	0
	Intr	65845	65980	+	0
25	Intr	66115	66290	+	0
	Term	66365	66942	+	0

>2564049 /37294

len = 1427 nex = 2

30	Term	485	294	-	0
	Init	1720	625	-	0

>2564049 /104793

len = 1873 nex = 7

	Init	17973	18128	+	0
	Intr	18663	18789	+	0
40	Intr	18882	19035	+	0
	Intr	19112	19208	+	0
	Intr	19304	19392	+	0
	Intr	19521	19589	+	0
45	Term	19790	19845	+	0

>2564049 /141731

len = 2068 nex = 7

50	Init	18007	18128	+	0
	Intr	18663	18789	+	0
	Intr	18882	19035	+	0
	Intr	19112	19208	+	0
	Intr	19304	19392	+	0
55	Intr	19521	19589	+	0
	Term	19790	20074	+	0

>2564049 /21604

60 len = 557 nex = 1

981

	Sngl	28618	28062	-	0
	>2564049	/16144			
5	len =	1365	nex =	3	
	Init	28919	29348	+	0
	Intr	29603	29695	+	0
10	Term	30029	30283	+	0
	>2564049	/31971			
15	len =	1818	nex =	2	
	Init	35677	36089	+	0
	Term	36890	37494	+	0
	>2564049	/13667			
20	len =	1704	nex =	6	
	Term	5026	4812	-	0
	Intr	5207	5118	-	0
25	Intr	5466	5299	-	0
	Intr	5691	5572	-	0
	Intr	5932	5787	-	0
	Init	6515	6354	-	0
30	>2564050	/6203			
	len =	2839	nex =	13	
	Init	12017	12391	+	0
35	Intr	12485	12567	+	0
	Intr	12820	12974	+	0
	Intr	13048	13082	+	0
	Intr	13144	13293	+	0
40	Intr	13467	13562	+	0
	Intr	13634	13750	+	0
	Intr	13832	13951	+	0
	Intr	14029	14121	+	0
	Intr	14202	14324	+	0
45	Intr	14407	14523	+	0
	Intr	14606	14668	+	0
	Term	14766	14842	+	0
	>2564050	/123496			
50	len =	674	nex =	1	
	Sngl	17696	18369	+	0
	>2564050	/16313			
55	len =	1594	nex =	5	
	Init	2671	2918	+	0
	Intr	3227	3325	+	0
60	Intr	3410	3518	+	0

005326-001600

982

	Intr	3687	3758	+	0
	Term	3993	4264	+	0
	>2564050 /14738				
5	len =	1040	nex =	2	
	Term	28103	27853	-	0
	Init	28892	28606	-	0
10	>2564050 /13951				
	len =	1063	nex =	2	
15	Term	28103	27848	-	0
	Init	28910	28606	-	0
	>2564050 /38057				
20	len =	2006	nex =	0	
	>2564051 /7688				
	len =	1722	nex =	2	
25	Term	13311	12928	-	0
	Init	13996	13887	-	0
	>2564051 /6220				
30	len =	2570	nex =	6	
	Init	18254	18493	+	0
	Intr	18575	18754	+	0
35	Intr	19785	19904	+	0
	Intr	19917	20078	+	0
	Intr	20178	20459	+	0
	Term	20546	20823	+	0
40	>2564051 /30648				
	len =	2334	nex =	9	
	Init	33401	33589	+	0
45	Intr	33676	33848	+	0
	Intr	34149	34268	+	0
	Intr	34373	34429	+	0
	Intr	34595	34675	+	0
	Intr	34763	34797	+	0
50	Intr	34933	35006	+	0
	Intr	35103	35262	+	0
	Term	35380	35734	+	0
	>2564051 /30994				
55	len =	2530	nex =	8	
	Init	45513	45608	+	0
	Intr	46036	46115	+	0
60	Intr	46206	46280	+	0

2025 RELEASE UNDER E.O. 14176

983

	Intr	46370	46473	+	0
	Intr	46561	46717	+	0
	Intr	46810	46897	+	0
	Intr	46997	47069	+	0
5	Term	47147	47224	+	0
	>2564051 /29619				
10	len =	942	nex =	3	
	Init	46810	46897	+	0
	Intr	46997	47069	+	0
	Term	47147	47224	+	0
15	>2564051 /29829				
	len =	1317	nex =	3	
	Term	48114	47710	-	0
20	Intr	48493	48207	-	0
	Init	49026	48809	-	0
	>2564051 /6519				
25	len =	1128	nex =	2	
	Init	72721	72978	+	0
	Term	73194	73848	+	0
30	>2564051 /142033				
	len =	651	nex =	2	
	Init	72788	72978	+	0
35	Term	73194	73438	+	0
	>2564051 /14159				
40	len =	1394	nex =	5	
	Term	74311	74056	-	0
	Intr	74603	74398	-	0
	Intr	74863	74713	-	0
	Intr	75172	74950	-	0
45	Init	75449	75412	-	0
	>2564051 /40866				
50	len =	1519	nex =	5	
	Term	74311	74064	-	0
	Intr	74603	74398	-	0
	Intr	74863	74713	-	0
	Intr	75172	74950	-	0
55	Init	75582	75412	-	0
	>2564051 /17770				
60	len =	1500	nex =	5	

2025-09-09 10:09:09

984

	Term	74311	74086	-	0
	Intr	74603	74398	-	0
	Intr	74863	74713	-	0
	Intr	75172	74950	-	0
5	Init	75445	75412	-	0
	>2564051 /13949				
10	len =	1110	nex =	3	
	Term	82879	82476	-	0
	Intr	83240	82973	-	0
	Init	83585	83325	-	0
15	>2570223 /40832				
	len =	2253	nex =	6	
20	Init	17162	17477	+	0
	Intr	17799	17892	+	0
	Intr	18430	18609	+	0
	Intr	18688	18807	+	0
	Intr	18887	19020	+	0
25	Term	19185	19414	+	0
	>2570223 /37699				
	len =	2869	nex =	9	
30	Term	26477	25979	-	0
	Intr	26840	26580	-	0
	Intr	27159	26941	-	0
	Intr	27498	27271	-	0
	Intr	27878	27776	-	0
35	Intr	28077	27965	-	0
	Intr	28258	28197	-	0
	Intr	28478	28346	-	0
	Init	28847	28757	-	0
40	>2570223 /23106				
	len =	629	nex =	1	
45	Sngl	74691	75319	+	0
	>2583106 /29207				
	len =	2272	nex =	5	
50	Init	108141	108430	+	0
	Intr	108875	109071	+	0
	Intr	109540	109629	+	0
	Intr	109744	109815	+	0
	Term	110152	110412	+	0
55	>2583106 /36389				
	len =	643	nex =	1	
60	Sngl	121587	120945	-	0

2025 RELEASE UNDER E.O. 14176

>2583106 /18320

5	len =	2568	nex =	7	
	Term	73308	72638	-	0
	Intr	73553	73400	-	0
	Intr	73796	73648	-	0
	Intr	74168	73886	-	0
10	Intr	74356	74251	-	0
	Intr	74536	74446	-	0
	Init	75205	74719	-	0

>2583106 /21765

15	len =	1874	nex =	0	
----	-------	------	-------	---	--

>2583106 /1969

20	len =	5689	nex =	1	
	Sngl	88355	88684	+	0

>2583106 /37127

25	len =	310	nex =	0	
----	-------	-----	-------	---	--

>2583106 /37621

30	len =	3101	nex =	15	
	Term	89198	88862	-	0
	Intr	89371	89306	-	0
	Intr	89531	89462	-	0
35	Intr	89689	89616	-	0
	Intr	89891	89793	-	0
	Intr	90037	89976	-	0
	Intr	90178	90137	-	0
	Intr	90316	90265	-	0
40	Intr	90541	90442	-	0
	Intr	90682	90638	-	0
	Intr	90843	90796	-	0
	Intr	91179	91104	-	0
	Intr	91456	91286	-	0
45	Intr	91590	91540	-	0
	Init	91962	91806	-	0

>2584827 /273

50	len =	2260	nex =	5	
	Term	101872	101586	-	0
	Intr	102093	102017	-	0
	Intr	102388	102242	-	0
55	Intr	102650	102480	-	0
	Init	103150	102928	-	0

>2584827 /5480

60	len =	1994	nex =	3	
----	-------	------	-------	---	--

	Term	111925	111310	-	0
	Intr	112167	112049	-	0
	Init	113303	112263	-	0
5	>2584827	/5171			
	len =	319	nex =	1	
10	Sngl	115114	114796	-	0
	>2584827	/17426			
	len =	597	nex =	1	
15	Sngl	115422	114826	-	0
	>2584827	/11593			
20	len =	562	nex =	1	
	Sngl	115422	114861	-	0
25	>2584827	/25571			
	len =	610	nex =	1	
	Sngl	115430	114821	-	0
30	>2584827	/34348			
	len =	1756	nex =	8	
35	Term	117147	116843	-	0
	Intr	117385	117233	-	0
	Intr	117590	117483	-	0
	Intr	117734	117687	-	0
	Intr	118025	117813	-	0
	Intr	118181	118117	-	0
40	Intr	118386	118262	-	0
	Init	118595	118482	-	0
	>2584827	/39107			
45	len =	2383	nex =	7	
	Term	117147	116840	-	0
	Intr	117385	117233	-	0
	Intr	117590	117483	-	0
50	Intr	117734	117687	-	0
	Intr	118025	117813	-	0
	Intr	118181	118117	-	0
	Init	118386	118262	-	0
55	>2584827	/5712			
	len =	790	nex =	1	
	Sngl	23900	24682	+	0
60					

2025 RELEASE UNDER E.O. 14176

	>2584827	/27675		
	len =	745	nex =	1
5	Sngl	23978	24722	+ 0
	>2584827	/116395		
	len =	286	nex =	1
10	Sngl	29868	29583	- 0
	>2584827	/4503		
	len =	471	nex =	1
15	Sngl	30113	29649	- 0
	>2584827	/22292		
	len =	592	nex =	1
20	Sngl	30233	29642	- 0
	>2584827	/25064		
	len =	683	nex =	2
	Term	30138	29660	- 0
30	Init	30342	30322	- 0
	>2584827	/25142		
	len =	816	nex =	2
	Term	30138	29596	- 0
35	Init	30411	30322	- 0
	>2584827	/1994		
	len =	835	nex =	2
	Term	30138	29584	- 0
40	Init	30418	30322	- 0
	>2584827	/4479		
	len =	655	nex =	3
	Term	84398	84092	- 0
50	Intr	84584	84486	- 0
	Init	84746	84674	- 0
	>2584827	/31676		
	len =	649	nex =	2
	Term	85483	85211	- 0
60	Init	85859	85582	- 0

2025 RELEASE UNDER E.O. 14176

>2584827 /32472

len = 1762 nex = 7

5	Term	84398	84113	-	0
	Intr	84584	84486	-	0
	Intr	84807	84674	-	0
	Intr	84997	84910	-	0
	Intr	85301	85255	-	0
10	Intr	85483	85417	-	0
	Init	85870	85582	-	0

>2584827 /8972

15 len = 4044 nex = 12

	Init	95183	95243	+	0
	Intr	95429	95523	+	0
	Intr	95608	95720	+	0
20	Intr	95804	95972	+	0
	Intr	96059	96098	+	0
	Intr	96231	96295	+	0
	Intr	96387	96500	+	0
	Intr	96601	96665	+	0
25	Intr	96783	96939	+	0
	Intr	97037	97156	+	0
	Intr	97247	97335	+	0
	Term	97422	97750	+	0

30 >2584827 /17473

len = 1881 nex = 9

	Init	95871	95972	+	0
35	Intr	96059	96098	+	0
	Intr	96231	96295	+	0
	Intr	96387	96500	+	0
	Intr	96601	96665	+	0
	Intr	96783	96939	+	0
40	Intr	97037	97156	+	0
	Intr	97247	97335	+	0
	Term	97422	97751	+	0

45 >2618599 /23293

len = 776 nex = 1

Sngl 11508 11343 - 0

50 >2618599 /6500

len = 997 nex = 1

Sngl 13043 14039 + 0

55 >2618599 /40212

len = 1750 nex = 3

60 Term 12611 12066 - 0

990

	Intr	13271	13086	-	0
	Init	13808	13354	-	0
	>2618599	/39514			
5	len =	1757	nex =	4	
	Term	23461	23198	-	0
	Intr	24300	24198	-	0
10	Intr	24625	24565	-	0
	Init	24954	24707	-	0
	>2618599	/8490			
15	len =	1570	nex =	5	
	Term	27344	27150	-	0
	Intr	27528	27425	-	0
	Intr	27900	27615	-	0
20	Intr	28171	27989	-	0
	Init	28711	28527	-	0
	>2618599	/96			
25	len =	1631	nex =	5	
	Term	27344	27085	-	0
	Intr	27528	27425	-	0
	Intr	27900	27615	-	0
30	Intr	28171	27989	-	0
	Init	28715	28527	-	0
	>2618599	/96124			
35	len =	490	nex =	1	
	Sngl	61246	60764	-	0
	>2618599	/13096			
40	len =	1229	nex =	3	
	Init	71559	71780	+	0
	Intr	72273	72395	+	0
45	Term	72494	72787	+	0
	>2618599	/93205			
50	len =	977	nex =	3	
	Init	71649	71780	+	0
	Intr	72273	72395	+	0
	Term	72494	72625	+	0
55	>2618599	/34629			
	len =	1114	nex =	3	
	Init	71649	71780	+	0
60	Intr	72273	72395	+	0

991

	Term	72494	72762	+	0
	>2618599 /29341				
5	len =	1153	nex =	3	
	Init	71650	71780	+	0
	Intr	72273	72395	+	0
	Term	72494	72802	+	0
10	>2618599 /40708				
	len =	1885	nex =	5	
15	Term	73223	72970	-	0
	Intr	73417	73329	-	0
	Intr	74198	74135	-	0
	Intr	74521	74458	-	0
	Init	74712	74609	-	0
20	>2618599 /37431				
	len =	2030	nex =	9	
25	Term	77559	77489	-	0
	Intr	77742	77644	-	0
	Intr	78021	77830	-	0
	Intr	78386	78249	-	0
	Intr	78575	78489	-	0
30	Intr	78745	78659	-	0
	Intr	79060	78899	-	0
	Intr	79293	79157	-	0
	Init	79506	79391	-	0
35	>2618600 /33948				
	len =	508	nex =	1	
	Sngl	11449	11709	+	0
40	>2618600 /5511				
	len =	2590	nex =	6	
45	Term	24538	24373	-	0
	Intr	24781	24651	-	0
	Intr	24969	24879	-	0
	Intr	25370	25228	-	0
	Intr	25905	25465	-	0
50	Init	26961	26323	-	0
	>2618600 /28326				
	len =	1413	nex =	2	
55	Term	34043	33702	-	0
	Init	35114	34932	-	0
60	>2618600 /25219				

2025 RELEASE UNDER E.O. 14176

992

	len =	1534	nex =	3	
	Init	36892	37172	+	0
	Intr	37302	37699	+	0
5	Term	37781	38425	+	0
	>2618600 /16323				
	len =	564	nex =	1	
10	Sngl	38835	39398	+	0
	>2618600 /19433				
	len =	3202	nex =	3	
	Init	54780	54996	+	0
	Intr	55178	55275	+	0
	Term	55405	57379	+	0
20	>2618600 /17022				
	len =	1656	nex =	0	
25	>2618600 /34218				
	len =	2898	nex =	6	
	Init	70436	70933	+	0
	Intr	71361	71426	+	0
	Intr	71892	72011	+	0
	Intr	72098	72382	+	0
	Intr	72467	72739	+	0
	Term	72924	73333	+	0
35	>2618600 /40280				
	len =	1001	nex =	3	
40	Term	80824	80546	-	0
	Intr	81167	81118	-	0
	Init	81369	81257	-	0
	>2618601 /98459				
45	len =	1457	nex =	3	
	Term	19035	18675	-	0
	Intr	19433	19407	-	0
50	Init	20131	19678	-	0
	>2618601 /38593				
	len =	1416	nex =	3	
55	Term	19035	18762	-	0
	Intr	19433	19407	-	0
	Init	20177	19678	-	0
60	>2618601 /3459				

2025-02-27 09:00:00

	len =	1481	nex =	3	
	Term	24639	24049	-	0
5	Intr	24900	24741	-	0
	Init	25529	25043	-	0
	>2618601	/41015			
10	len =	90	nex =	1	
	Sngl	2878	2789	-	0
	>2618601	/2577			
15	len =	1537	nex =	3	
	Term	28396	27900	-	0
	Intr	28801	28642	-	0
20	Init	29436	28936	-	0
	>2618601	/23349			
25	len =	1680	nex =	3	
	Term	33360	32729	-	0
	Intr	33823	33661	-	0
	Init	34408	33933	-	0
30	>2618601	/10032			
	len =	1835	nex =	5	
	Term	37993	37806	-	0
35	Intr	38138	38078	-	0
	Intr	38275	38222	-	0
	Intr	38435	38388	-	0
	Init	38573	38511	-	0
40	>2618601	/28362			
	len =	876	nex =	2	
	Term	41130	40811	-	0
45	Init	41686	41620	-	0
	>2618601	/36501			
	len =	970	nex =	2	
50	Term	41130	40731	-	0
	Init	41698	41620	-	0
	>2618601	/31107			
55	len =	1531	nex =	4	
	Init	44336	44550	+	0
	Intr	44852	44969	+	0
60	Intr	45277	45424	+	0

20250722-22252600

994

	Term	45511	45866	+	0
	>2618602	/16835			
5	len =	1213	nex =	3	
	Init	1864	2123	+	0
	Intr	2566	2722	+	0
	Term	2846	3076	+	0
10	>2618602	/30658			
	len =	1570	nex =	5	
15	Term	22604	22192	-	0
	Intr	22912	22693	-	0
	Intr	23242	23001	-	0
	Intr	23527	23315	-	0
	Init	23761	23601	-	0
20	>2618602	/37357			
	len =	1215	nex =	1	
25	Sngl	28713	29927	+	0
	>2618602	/25423			
	len =	677	nex =	1	
30	Sngl	46717	47393	+	0
	>2618602	/99873			
35	len =	900	nex =	1	
	Sngl	54255	53356	-	0
40	>2618602	/7672			
	len =	267	nex =	1	
	Sngl	59695	59434	-	0
45	>2618602	/10243			
	len =	1424	nex =	5	
	Term	59695	59432	-	0
50	Intr	59897	59845	-	0
	Intr	60072	60011	-	0
	Intr	60287	60254	-	0
	Init	60855	60687	-	0
55	>2618602	/1888			
	len =	1916	nex =	4	
	Init	66033	66354	+	0
60	Intr	66640	66827	+	0

995

Intr	66906	67302	+	0
Term	67408	67948	+	0

```

5      >2618602      /3109
      len =      4196      nex =      5
      Term      76502      75640      -      0
10     Intr      76837      76585      -      0
      Intr      77507      76944      -      0
      Intr      78819      78722      -      0
      Init      79835      79282      -      0

```

15	>2618602	/43035			
	len =	2160	nex =	8	
	Term	6638	6237	-	0
	Intr	6992	6725	-	0
20	Intr	7269	7075	-	0
	Intr	7586	7446	-	0
	Intr	7831	7765	-	0
	Intr	8009	7930	-	0
	Intr	8166	8088	-	0
25	Init	8396	8321	-	0

	>2618602	/36948			
30	len =	2038	nex =	8	
	Term	6638	6378	-	0
	Intr	6992	6725	-	0
	Intr	7269	7075	-	0
35	Intr	7586	7446	-	0
	Intr	7831	7765	-	0
	Intr	8009	7930	-	0
	Intr	8166	8088	-	0
	Init	8415	8321	-	0

```

40  >2618602      /22582
      len =      1527  nex =      5
45  Term      7586      7446      -      0
     Intr     7831      7765      -      0
     Intr     8009      7930      -      0
     Intr     8166      8088      -      0
     Init     8415      8321      -      0

```

```
50  >2618603          /143299
      len =      678  nex =      1
      Snq1 13660 12983      -      0
```

```

55 >2618603          /17250
    len =      2810  nex =          7
60   Term 22323  21890      -      0

```

996

	Intr	22694	22411	-	0
	Intr	22944	22785	-	0
	Intr	23379	23050	-	0
	Intr	23590	23464	-	0
5	Intr	23755	23676	-	0
	Init	24699	24401	-	0

>2618603 /40850

10	len =	2893	nex =	13	
	Term	57315	57005	-	0
	Intr	57504	57416	-	0
	Intr	57696	57585	-	0
15	Intr	57938	57801	-	0
	Intr	58165	58019	-	0
	Intr	58344	58270	-	0
	Intr	58476	58435	-	0
	Intr	58720	58601	-	0
20	Intr	58890	58816	-	0
	Intr	59036	58974	-	0
	Intr	59195	59126	-	0
	Intr	59447	59288	-	0
25	Init	59897	59584	-	0

>2618604 /17002

	len =	2138	nex =	7	
30	Term	48168	47893	-	0
	Intr	48328	48286	-	0
	Intr	48523	48450	-	0
	Intr	48712	48639	-	0
	Intr	49301	49226	-	0
35	Intr	49476	49399	-	0
	Init	50030	49757	-	0

>2618604 /32773

40	len =	419	nex =	1	
	Sngl	50632	50223	-	0

>2618605 /34592

45	len =	2890	nex =	10	
	Term	16301	16104	-	0
	Intr	16453	16385	-	0
50	Intr	16645	16540	-	0
	Intr	16810	16755	-	0
	Intr	17032	16973	-	0
	Intr	17226	17142	-	0
	Intr	17695	17601	-	0
55	Intr	17967	17845	-	0
	Intr	18278	18215	-	0
	Init	18991	18733	-	0

>2618605 /15525

60

20250720 09:13:53

997

	len =	1379	nex =	3	
	Init	19682	19907	+	0
	Intr	20580	20660	+	0
5	Term	20864	21060	+	0
	>2618605	/18394			
	len =	924	nex =	0	
10	>2618605	/37433			
	len =	1210	nex =	2	
	Init	19695	19907	+	0
15	Term	20580	20660	+	0
	>2618605	/98771			
	len =	1392	nex =	3	
	Init	19695	19907	+	0
	Intr	20580	20660	+	0
	Term	20864	21086	+	0
20	>2618605	/4869			
	len =	1372	nex =	0	
	>2618605	/6170			
	len =	1030	nex =	2	
	Term	21605	21167	-	0
	Init	22188	21736	-	0
25	>2618605	/114037			
	len =	370	nex =	1	
	Sngl	29570	29939	+	0
	>2618605	/39018			
	len =	1334	nex =	0	
	>2618605	/12152			
	len =	1346	nex =	0	
	>2618605	/11651			
	len =	1166	nex =	1	
	Sngl	47145	48310	+	0
	>2618605	/8440			
	len =	1831	nex =	7	

998

	Term	55825	55588	-	0
	Intr	56233	56081	-	0
	Intr	56458	56347	-	0
	Intr	56617	56546	-	0
5	Intr	56796	56703	-	0
	Intr	57030	56879	-	0
	Init	57418	57264	-	0
	>2618677		/14480		
10	len =	1296	nex =	3	
	Term	12884	12633	-	0
	Intr	13526	13383	-	0
15	Init	13928	13599	-	0
	>2618677		/7573		
	len =	1708	nex =	5	
20	Term	28265	27997	-	0
	Intr	28700	28345	-	0
	Intr	29322	28785	-	0
	Intr	29506	29408	-	0
25	Init	29704	29600	-	0
	>2618677		/111449		
	len =	748	nex =	1	
30	Sngl	38479	37732	-	0
	>2618677		/30006		
35	len =	2334	nex =	12	
	Init	49398	49629	+	0
	Intr	49719	49823	+	0
	Intr	49924	49986	+	0
40	Intr	50096	50129	+	0
	Intr	50217	50339	+	0
	Intr	50414	50476	+	0
	Intr	50565	50658	+	0
	Intr	50755	50800	+	0
45	Intr	50894	50961	+	0
	Intr	51110	51158	+	0
	Intr	51254	51319	+	0
	Term	51417	51731	+	0
50	>2618677		/30716		
	len =	804	nex =	1	
	Sngl	5723	6526	+	0
55	>2618677		/29451		
	len =	850	nex =	1	
60	Sngl	61457	62306	+	0

2025 RELEASE UNDER E.O. 14176

	>2618677	/141763		
5	len =	915	nex =	1
	Sngl	6460	5546	- 0
	>2618677	/18592		
10	len =	342	nex =	1
	Sngl	76241	76582	+ 0
	>2618677	/452		
15	len =	1630	nex =	6
	Term	88176	87859	- 0
	Intr	88357	88285	- 0
20	Intr	88547	88476	- 0
	Intr	88713	88642	- 0
	Intr	88988	88858	- 0
	Init	89480	89128	- 0
25	>2618677	/5532		
	len =	915	nex =	1
	Sngl	89616	90530	+ 0
30	>2618683	/37276		
	len =	1974	nex =	7
35	Term	24726	24402	- 0
	Intr	24881	24816	- 0
	Intr	25061	24963	- 0
	Intr	25435	25239	- 0
	Intr	25623	25576	- 0
40	Intr	25828	25711	- 0
	Init	26375	26175	- 0
	>2618683	/40508		
45	len =	2051	nex =	7
	Term	24726	24397	- 0
	Intr	24881	24816	- 0
	Intr	25061	24963	- 0
50	Intr	25435	25239	- 0
	Intr	25623	25576	- 0
	Intr	25828	25711	- 0
	Init	26447	26175	- 0
55	>2618683	/13347		
	len =	2662	nex =	6
	Init	27061	27168	+ 0
60	Intr	27764	28346	+ 0

1000

	Intr	28448	28669	+	0
	Intr	28748	28995	+	0
	Intr	29088	29227	+	0
5	Term	29422	29722	+	0
	>2618683	/2657			
	len =	2247	nex =	7	
10	Term	56579	56243	-	0
	Intr	56748	56672	-	0
	Intr	56919	56831	-	0
	Intr	57186	57056	-	0
	Intr	57409	57280	-	0
15	Intr	57542	57518	-	0
	Init	58489	57939	-	0
	>2618683	/2085			
20	len =	561	nex =	2	
	Init	64780	64842	+	0
	Term	64924	65340	+	0
25	>2618683	/35456			
	len =	1030	nex =	1	
	Sngl	68721	69748	+	0
30	>2623294	/6610			
	len =	1969	nex =	2	
35	Init	13412	14121	+	0
	Term	14135	14416	+	0
	>2623294	/11739			
40	len =	521	nex =	2	
	Init	27204	27453	+	0
	Term	27544	27724	+	0
45	>2623294	/1297			
	len =	1365	nex =	3	
	Init	52988	53175	+	0
50	Intr	53533	53759	+	0
	Term	53848	54135	+	0
	>2623294	/2337			
55	len =	4786	nex =	4	
	Init	65365	65544	+	0
	Intr	65952	66570	+	0
	Intr	66666	66939	+	0
60	Term	67114	67245	+	0

1001

>2623294 /30905

5	len =	984	nex =	2	
	Init	71769	71892	+	0
	Term	72257	72752	+	0

>2642152 /105724

10	len =	1015	nex =	0	
----	-------	------	-------	---	--

>2642152 /5605

15	len =	1069	nex =	1	
	Sngl	12014	12116	+	0

>2642152 /123288

20	len =	600	nex =	2	
	Term	13599	13304	-	0
	Init	13903	13724	-	0

>2642152 /18133

25	len =	271	nex =	1	
30	Sngl	21652	21382	-	0

>2642152 /8256

35	len =	3019	nex =	4	
	Init	25623	26134	+	0
	Intr	26620	26968	+	0
	Intr	27055	27331	+	0
	Term	28305	28641	+	0

>2642152 /37775

40	len =	2269	nex =	4	
45	Init	31242	31353	+	0
	Intr	32126	32188	+	0
	Intr	32277	32516	+	0
	Term	32603	33510	+	0

>2642152 /111626

50	len =	250	nex =	1	
	Sngl	33270	33510	+	0

>2642152 /27496

55	len =	2548	nex =	9	
60	Init	33756	33970	+	0

20250910-09123563

1002

	Intr	34189	34217	+	0
	Intr	34335	34408	+	0
	Intr	34530	34631	+	0
	Intr	34670	34777	+	0
5	Intr	35009	35100	+	0
	Intr	35365	35526	+	0
	Intr	35796	35880	+	0
	Term	35999	36303	+	0
10	>2642152		/39203		
	len =	1168	nex =	3	
	Init	39520	39810	+	0
15	Term	39958	40298	+	0
	Term	40402	40687	+	0
	>2642152		/17221		
20	len =	2364	nex =	13	
	Term	41130	41068	-	0
	Intr	41304	41234	-	0
	Intr	41463	41397	-	0
25	Intr	41617	41549	-	0
	Intr	41866	41716	-	0
	Intr	42123	42048	-	0
	Intr	42323	42220	-	0
	Intr	42505	42409	-	0
30	Intr	42680	42594	-	0
	Intr	42849	42763	-	0
	Intr	43018	42929	-	0
	Intr	43229	43152	-	0
	Init	43431	43342	-	0
35	>2642152		/158481		
	len =	2590	nex =	6	
40	Term	52943	52448	-	0
	Intr	53435	53162	-	0
	Intr	53934	53766	-	0
	Intr	54202	54122	-	0
45	Intr	54359	54302	-	0
	Init	55032	54833	-	0
	>2642152		/114221		
	len =	321	nex =	1	
50	Sngl	69754	70074	+	0
	>2642427		/28853		
55	len =	1030	nex =	2	
	Term	106334	106067	-	0
	Init	107087	106553	-	0
60	>2642427		/30636		

1003

	len =	1167	nex =	3	
5	Term	25457	25117	-	0
	Intr	25685	25560	-	0
	Init	26283	25787	-	0
	>2642427 /38686				
10	len =	1419	nex =	3	
	Term	25457	25104	-	0
	Intr	25685	25560	-	0
	Init	26522	25787	-	0
15	>2642427 /25204				
	len =	1692	nex =	3	
20	Term	32565	32226	-	0
	Intr	32850	32670	-	0
	Init	33917	33379	-	0
	>2642427 /40307				
25	len =	1118	nex =	1	
	Sngl	45533	46650	+	0
30	>2642427 /40943				
	len =	631	nex =	1	
	Sngl	48222	47592	-	0
35	>2642427 /22551				
	len =	670	nex =	2	
40	Term	5656	5170	-	0
	Init	5836	5750	-	0
	>2642427 /22936				
45	len =	3101	nex =	9	
	Term	5656	5284	-	0
	Intr	5932	5750	-	0
	Intr	6183	6016	-	0
50	Intr	6419	6285	-	0
	Intr	6590	6501	-	0
	Intr	6761	6684	-	0
	Intr	7515	6859	-	0
	Intr	7718	7600	-	0
55	Init	8384	7931	-	0
	>2645198 /28033				
60	len =	817	nex =	1	

1004

	Sngl	15766	16582	+	0
	>2645198	/144393			
5	len =	651	nex =	2	
	Term	43394	43065	-	0
	Init	43715	43483	-	0
10	>2645198	/34660			
	len =	676	nex =	2	
	Term	51571	51243	-	0
15	Init	51918	51697	-	0
	>2645198	/21053			
	len =	691	nex =	3	
20	Term	54371	54189	-	0
	Intr	54602	54463	-	0
	Init	54879	54688	-	0
25	>2645198	/36567			
	len =	1873	nex =	2	
	Term	6681	6341	-	0
30	Init	8213	7815	-	0
	>2651294	/6884			
	len =	1979	nex =	8	
35	Init	119	467	+	0
	Intr	563	788	+	0
	Intr	984	1065	+	0
	Intr	1166	1223	+	0
40	Intr	1300	1388	+	0
	Intr	1482	1577	+	0
	Intr	1671	1693	+	0
	Term	1814	2097	+	0
45	>2651294	/99409			
	len =	1607	nex =	3	
	Init	12703	12929	+	0
50	Intr	13029	13341	+	0
	Term	13833	14309	+	0
	>2651294	/19516			
55	len =	1380	nex =	4	
	Init	17001	17093	+	0
	Intr	17192	17276	+	0
	Intr	17399	17532	+	0
60	Term	17965	18220	+	0

2025 RELEASE UNDER E.O. 14176

1005

>2651294 /25574

5	len =	610	nex =	1	
	Sngl	29291	28700	-	0

>2651294 /15024

10	len =	695	nex =	1	
	Sngl	35758	35064	-	0

>2651294 /28121

15	len =	1330	nex =	4	
	Term	43723	43497	-	0
	Intr	44224	44091	-	0
20	Intr	44474	44390	-	0
	Init	44667	44575	-	0

>2651294 /120147

25	len =	792	nex =	1	
	Sngl	48787	48117	-	0

>2651294 /12956

30	len =	1078	nex =	1	
	Sngl	55272	56349	+	0

>2651294 /97343

35	len =	1233	nex =	3	
	Init	59647	59727	+	0
40	Intr	59886	60144	+	0
	Term	60429	60879	+	0

>2651294 /35274

45	len =	3115	nex =	15	
	Init	80840	81063	+	0
	Intr	81295	81464	+	0
	Intr	81548	81599	+	0
50	Intr	81692	81717	+	0
	Intr	81829	81912	+	0
	Intr	82014	82124	+	0
	Intr	82222	82304	+	0
	Intr	82425	82515	+	0
55	Intr	82617	82668	+	0
	Intr	82768	82858	+	0
	Intr	82979	83101	+	0
	Intr	83207	83329	+	0
	Intr	83465	83531	+	0
60	Intr	83626	83705	+	0

20250120 09:29:56

1006

	Term	83788	83954	+	0
	>2656024	/2925			
5	len =	2110	nex =	6	
	Init	21471	21726	+	0
	Intr	21825	21900	+	0
	Intr	21980	22075	+	0
10	Intr	22220	22557	+	0
	Intr	23046	23175	+	0
	Term	23299	23578	+	0
	>2656024	/10982			
15	len =	2112	nex =	8	
	Term	36311	35994	-	0
	Intr	36483	36427	-	0
20	Intr	36652	36568	-	0
	Intr	36876	36751	-	0
	Intr	37003	36960	-	0
	Intr	37209	37090	-	0
	Intr	37445	37291	-	0
25	Init	38105	38035	-	0
	>2656024	/115913			
	len =	454	nex =	1	
30	Sngl	69260	69713	+	0
	>2656025	/42551			
35	len =	1597	nex =	3	
	Init	21954	22111	+	0
	Intr	22318	22630	+	0
	Term	22710	23356	+	0
40	>2656025	/17690			
	len =	2548	nex =	6	
45	Init	26822	27289	+	0
	Intr	27502	27719	+	0
	Intr	27833	27945	+	0
	Intr	28683	28746	+	0
	Intr	28894	28947	+	0
50	Term	29171	29369	+	0
	>2656025	/31580			
	len =	2657	nex =	3	
55	Term	34185	33154	-	0
	Intr	34744	34283	-	0
	Init	35810	35427	-	0
60	>2656025	/21044			

1007

	len =	4643	nex =	16	
	Term	36770	36474	-	0
5	Intr	36904	36853	-	0
	Intr	37063	37002	-	0
	Intr	37298	37211	-	0
	Intr	37535	37467	-	0
	Intr	37706	37649	-	0
10	Intr	37887	37786	-	0
	Intr	38077	37976	-	0
	Intr	38499	38169	-	0
	Intr	38817	38701	-	0
	Intr	38980	38912	-	0
15	Intr	39610	39366	-	0
	Intr	40211	39972	-	0
	Intr	40458	40363	-	0
	Intr	40839	40723	-	0
20	Init	41116	40940	-	0
	>2656026	/27668			
	len =	641	nex =	1	
25	Sngl	21786	22426	+	0
	>2656026	/32253			
	len =	1708	nex =	2	
30	Init	22982	24225	+	0
	Term	24310	24689	+	0
	>2656026	/39882			
35	len =	1499	nex =	2	
	Init	22997	24225	+	0
	Term	24310	24471	+	0
40	>2656026	/205741			
	len =	1213	nex =	4	
45	Init	33785	34004	+	0
	Intr	34198	34304	+	0
	Intr	34400	34430	+	0
	Term	34537	34990	+	0
50	>2656026	/27452			
	len =	1064	nex =	1	
	Sngl	39160	40223	+	0
55	>2656026	/34082			
	len =	773	nex =	1	
60	Sngl	47232	46460	-	0

1008

	>2656026	/42826		
5	len =	1690	nex =	3
	Init	47575	48443	+
	Intr	48536	48674	+
	Term	48854	49260	+
10	>2656026	/9664		
	len =	810	nex =	1
15	Sngl	50539	49730	-
	>2656026	/16760		
	len =	3571	nex =	0
20	>2656027	/37474		
	len =	5143	nex =	4
25	Init	22714	22946	+
	Intr	23554	23642	+
	Intr	23726	23824	+
	Term	23923	24012	+
	>2656027	/117516		
30	len =	1426	nex =	4
	Init	22758	22946	+
	Intr	23554	23642	+
35	Intr	23726	23824	+
	Term	23923	24183	+
	>2656027	/96557		
40	len =	1396	nex =	4
	Init	22776	22946	+
	Intr	23554	23642	+
	Intr	23726	23824	+
45	Term	23923	24171	+
	>2656027	/34532		
	len =	2155	nex =	4
50	Init	24706	25078	+
	Intr	25689	26087	+
	Intr	26183	26517	+
	Term	26605	26860	+
55	>2656028	/39718		
	len =	910	nex =	2
60	Init	21537	21782	+

1009

	Term	22349	22446	+	0
	>2656028 /11128				
5	len =	1591	nex =	5	
	Init	21555	21782	+	0
	Intr	22349	22446	+	0
	Intr	22555	22722	+	0
10	Intr	22802	22844	+	0
	Term	22931	23145	+	0
	>2656028 /38185				
15	len =	2448	nex =	8	
	Term	23351	23070	-	0
	Intr	23563	23476	-	0
	Intr	23838	23740	-	0
20	Intr	24026	23922	-	0
	Intr	24195	24126	-	0
	Intr	24373	24277	-	0
	Intr	24640	24499	-	0
	Init	25228	25069	-	0
25	>2656028 /27013				
	len =	2579	nex =	9	
30	Term	23351	23054	-	0
	Intr	23563	23476	-	0
	Intr	23838	23740	-	0
	Intr	24026	23922	-	0
	Intr	24195	24126	-	0
35	Intr	24373	24277	-	0
	Intr	24640	24499	-	0
	Intr	25228	25069	-	0
	Init	25632	25564	-	0
40	>2656028 /28596				
	len =	2454	nex =	10	
	Term	39978	39624	-	0
45	Intr	40161	40113	-	0
	Intr	40298	40244	-	0
	Intr	40511	40391	-	0
	Intr	40705	40604	-	0
	Intr	40916	40809	-	0
50	Intr	41213	40996	-	0
	Intr	41361	41289	-	0
	Intr	41523	41447	-	0
	Init	42077	41767	-	0
	>2656028 /103642				
55	len =	690	nex =	1	
	Sngl	55512	55324	-	0
60	>2656028 /36094				

1010

	len =	2124	nex =	3	
5	Init	56165	56791	+	0
	Intr	56822	56893	+	0
	Term	57886	58288	+	0
	>2656028 /3303				
10	len =	1667	nex =	4	
	Term	58518	58274	-	0
	Intr	58878	58601	-	0
	Intr	59124	58950	-	0
15	Init	59940	59506	-	0
	>2656028 /18435				
20	len =	2060	nex =	8	
	Init	60472	60847	+	0
	Intr	60939	61071	+	0
	Intr	61351	61440	+	0
	Intr	61533	61657	+	0
25	Intr	61748	61790	+	0
	Intr	61872	61989	+	0
	Intr	62089	62138	+	0
	Term	62218	62531	+	0
30	>2656028 /11539				
	len =	864	nex =	1	
35	Sngl	65472	66335	+	0
	>2656028 /34900				
	len =	430	nex =	1	
40	Sngl	75838	76262	+	0
	>2656028 /16977				
45	len =	430	nex =	1	
	Sngl	75840	76262	+	0
	>2656028 /155377				
50	len =	850	nex =	5	
	Init	9258	9512	+	0
	Intr	9598	9659	+	0
	Intr	9739	9790	+	0
55	Intr	9876	9990	+	0
	Term	10076	10107	+	0
	>2656029 /2641				
60	len =	1789	nex =	5	

1011

	Init	15938	16144	+	0
	Intr	16242	16362	+	0
	Intr	16458	16549	+	0
5	Intr	16632	16728	+	0
	Term	16810	17371	+	0
	>2656029 /1843				
10	len =	670	nex =	1	
	Sngl	18752	18085	-	0
	>2656029 /578				
15	len =	1630	nex =	7	
	Init	36717	36759	+	0
	Intr	36898	36971	+	0
20	Intr	37133	37178	+	0
	Intr	37268	37338	+	0
	Intr	37454	37498	+	0
	Intr	37595	37663	+	0
	Term	37791	38345	+	0
25	>2656029 /22827				
	len =	884	nex =	1	
30	Sngl	374	1257	+	0
	>2656029 /5589				
35	len =	1459	nex =	3	
	Term	43267	42861	-	0
	Intr	43429	43348	-	0
	Init	43780	43535	-	0
40	>2656029 /30314				
	len =	915	nex =	1	
45	Sngl	48635	49549	+	0
	>2656029 /38273				
	len =	3043	nex =	4	
50	Init	6585	6905	+	0
	Intr	7927	8060	+	0
	Intr	8561	8708	+	0
	Term	9213	9319	+	0
55	>2656029 /10991				
	len =	550	nex =	1	
60	Sngl	66286	66832	+	0

1012

>2656030 /42911

len = 933 nex = 3

5	Init	78500	78616	+	0
	Intr	78721	78891	+	0
	Term	79064	79432	+	0

>2656030 /5331

10	len =	1630	nex =	5	
	Term	79640	79440	-	0
	Intr	79934	79737	-	0
15	Intr	80170	80021	-	0
	Intr	80795	80709	-	0
	Init	81069	80880	-	0

>2656031 /6333

20	len =	1318	nex =	2	
	Init	18702	19086	+	0
	Term	19654	20019	+	0

>2656032 /31040

	len =	3269	nex =	10	
30	Init	17542	17641	+	0
	Intr	17886	18084	+	0
	Intr	18311	18413	+	0
	Intr	18505	18575	+	0
	Intr	18958	19119	+	0
35	Intr	19218	19355	+	0
	Intr	19444	19737	+	0
	Intr	19841	19993	+	0
	Intr	20110	20205	+	0
40	Term	20311	20555	+	0

>2656032 /1048

len = 1073 nex = 1

45	Sngl	34012	32940	-	0
----	------	-------	-------	---	---

>2660661 /31042

len = 1330 nex = 1

50	Sngl	12325	13650	+	0
----	------	-------	-------	---	---

>2660661 /1383

55	len =	1030	nex =	2	
----	-------	------	-------	---	--

	Term	64341	63993	-	0
	Init	65022	64911	-	0

60	>2660661	/39167			
----	----------	--------	--	--	--

2025-08-16 16:00:00

1013

	len =	1410	nex =	3	
5	Init	77572	77856	+	0
	Intr	77938	78144	+	0
	Term	78228	78610	+	0
	>2660661 /4029				
10	len =	2474	nex =	8	
	Term	78971	78641	-	0
	Intr	79182	79069	-	0
	Intr	79367	79265	-	0
15	Intr	79521	79457	-	0
	Intr	79694	79606	-	0
	Intr	79951	79775	-	0
	Intr	80178	80029	-	0
	Init	81114	80933	-	0
20	>2673901 /12267				
	len =	1283	nex =	1	
25	Sngl	21553	21782	+	0
	>2673901 /33002				
	len =	2397	nex =	3	
30	Init	25887	26600	+	0
	Intr	27248	27393	+	0
	Term	27470	28283	+	0
35	>2673901 /28982				
	len =	1630	nex =	2	
	Init	29263	29613	+	0
40	Term	30084	30468	+	0
	>2673901 /32748				
	len =	3159	nex =	9	
45	Term	32389	32054	-	0
	Intr	32585	32474	-	0
	Intr	32774	32682	-	0
	Intr	33085	32882	-	0
50	Intr	33504	33364	-	0
	Intr	33696	33592	-	0
	Intr	34178	33798	-	0
	Intr	34443	34303	-	0
	Init	35212	34776	-	0
55	>2673901 /38748				
	len =	235	nex =	1	
60	Sngl	35473	35707	+	0

1014

>2689438 /151406

5	len =	472	nex =	1	
	Sngl	16845	17316	+	0

>2689438 /3000

10	len =	2973	nex =	4	
	Term	25402	24983	-	0
	Intr	26284	26031	-	0
	Intr	27230	26975	-	0
15	Init	27948	27691	-	0

>2689438 /31290

20	len =	1945	nex =	4	
	Init	48660	48849	+	0
	Intr	48947	49173	+	0
	Intr	49650	49983	+	0
	Term	50177	50604	+	0

>2689438 /25158

	len =	448	nex =	1	
30	Sngl	50196	50643	+	0

>2689438 /35916

35	len =	2140	nex =	8	
	Term	72752	72356	-	0
	Intr	72927	72844	-	0
	Intr	73134	73078	-	0
	Intr	73359	73231	-	0
40	Intr	73775	73689	-	0
	Intr	73908	73861	-	0
	Intr	74175	74019	-	0
	Init	74495	74380	-	0

45 >2696018 /143114

	len =	250	nex =	1	
	Sngl	24191	24437	+	0

50 >2696018 /5698

	len =	1017	nex =	1	
55	Sngl	32527	32321	-	0

>2696018 /92795

60	len =	654	nex =	1	
----	-------	-----	-------	---	--

1015

	Sngl	49709	49056	-	0
--	------	-------	-------	---	---

	>2696018	/6629		
--	----------	-------	--	--

5	len =	2745	nex =	8	
---	-------	------	-------	---	--

	Init	56009	56062	+	0
--	------	-------	-------	---	---

	Intr	56175	56325	+	0
--	------	-------	-------	---	---

	Intr	56866	56932	+	0
--	------	-------	-------	---	---

10	Intr	57026	57174	+	0
----	------	-------	-------	---	---

	Intr	57438	57571	+	0
--	------	-------	-------	---	---

	Intr	57657	57755	+	0
--	------	-------	-------	---	---

	Intr	57849	57924	+	0
--	------	-------	-------	---	---

	Term	58005	58355	+	0
--	------	-------	-------	---	---

15	>2696018	/11275		
----	----------	--------	--	--

	len =	2084	nex =	9	
--	-------	------	-------	---	--

20	Init	56016	56062	+	0
----	------	-------	-------	---	---

	Intr	56175	56325	+	0
--	------	-------	-------	---	---

	Intr	56866	56932	+	0
--	------	-------	-------	---	---

	Intr	57026	57174	+	0
--	------	-------	-------	---	---

	Intr	57370	57405	+	0
--	------	-------	-------	---	---

25	Intr	57438	57571	+	0
----	------	-------	-------	---	---

	Intr	57657	57755	+	0
--	------	-------	-------	---	---

	Intr	57849	57924	+	0
--	------	-------	-------	---	---

	Term	58005	58099	+	0
--	------	-------	-------	---	---

30	>2696018	/4330		
----	----------	-------	--	--

	len =	1842	nex =	6	
--	-------	------	-------	---	--

	Term	67677	67445	-	0
--	------	-------	-------	---	---

35	Intr	67827	67768	-	0
----	------	-------	-------	---	---

	Intr	68176	68136	-	0
--	------	-------	-------	---	---

	Intr	68784	68715	-	0
--	------	-------	-------	---	---

	Intr	68912	68859	-	0
--	------	-------	-------	---	---

40	Init	69286	69237	-	0
----	------	-------	-------	---	---

	>2696018	/37111		
--	----------	--------	--	--

	len =	992	nex =	2	
--	-------	-----	-------	---	--

45	Term	79220	78599	-	0
----	------	-------	-------	---	---

	Init	79576	79309	-	0
--	------	-------	-------	---	---

	>2702261	/26812		
--	----------	--------	--	--

50	len =	2400	nex =	7	
----	-------	------	-------	---	--

	Init	2306	2536	+	0
--	------	------	------	---	---

	Intr	2938	2981	+	0
--	------	------	------	---	---

	Intr	3093	3250	+	0
--	------	------	------	---	---

55	Intr	3640	3767	+	0
----	------	------	------	---	---

	Intr	3879	4001	+	0
--	------	------	------	---	---

	Intr	4101	4220	+	0
--	------	------	------	---	---

	Term	4310	4705	+	0
--	------	------	------	---	---

60	>2702261	/3972		
----	----------	-------	--	--

1017

>2708736 /12534

5	len =	808	nex =	1	
	Sngl	57189	56382	-	0

>2708736 /24747

10	len =	1231	nex =	1	
	Sngl	59076	60306	+	0

>2708736 /9676

15	len =	895	nex =	4	
	Init	63519	63714	+	0
	Intr	63796	63961	+	0
20	Intr	64080	64133	+	0
	Term	64232	64413	+	0

>2708736 /40806

25	len =	2170	nex =	4	
	Init	66493	67471	+	0
	Intr	67805	67971	+	0
	Intr	68059	68143	+	0
30	Term	68241	68655	+	0

>2739359 /22595

35	len =	2809	nex =	5	
	Term	29529	28886	-	0
	Intr	30066	29793	-	0
	Intr	30458	30289	-	0
	Intr	30730	30590	-	0
40	Init	30933	30852	-	0

>2739359 /9368

45	len =	715	nex =	2	
	Init	55832	55932	+	0
	Term	56029	56546	+	0

>2739359 /30772

50	len =	673	nex =	1	
	Sngl	55861	55932	+	0

>2739359 /14528

55	len =	644	nex =	1	
	Sngl	56029	56527	+	0

60

1018

>2739359 /31774

```
len = 2950    nex = 12
```

		Term	3431	3158	-	0
5	Intr	3534	3459	-	0	
	Intr	3857	3781	-	0	
	Intr	4371	4277	-	0	
	Intr	4756	4623	-	0	
	Intr	4924	4846	-	0	
10	Intr	5064	4995	-	0	
	Intr	5348	5159	-	0	
	Intr	5522	5428	-	0	
	Intr	5694	5605	-	0	
15	Intr	5801	5770	-	0	
	Init	6098	5900	-	0	

>2739359 /628

```
20      len =      827      nex =      1
      Snq1 65777 66603      +      0
```

>2739359 /33680

```
len =      634    nex =      1
Snq1 69683 69050      -      0
```

30 >2739359 /1999

len = 1979 nex = 6

35	Term	79655	79490	-	0
	Intr	79924	79752	-	0
	Term	80307	80005	-	0
	Intr	80697	80536	-	0
	Term	81152	80784	-	0
	Intr	81468	81236	-	0

>2749918 /27548

```
len = 1572    nex = 4
```

45	Term	118910	118818	-	0
	Intr	119099	119049	-	0
	Intr	119342	119186	-	0
	Init	119799	119495	-	0

50 >2749918 /2157

```
len = 1164    nex = 5
```

55	Term	33097	32732	-	(
	Intr	33226	33170	-	(
	Intr	33418	33313	-	(
	Intr	33578	33520	-	(
	Init	33895	33656	-	(

60 >2749918 /41337

1019

	len =	2907	nex =	8	
	Term	40133	39852	-	0
5	Intr	40316	40251	-	0
	Intr	40452	40399	-	0
	Intr	40631	40560	-	0
	Intr	40820	40725	-	0
	Intr	41443	41366	-	0
10	Intr	41920	41841	-	0
	Init	42758	42004	-	0
	>2749918	/5055			
15	len =	2249	nex =	8	
	Term	73192	72956	-	0
	Intr	73427	73285	-	0
	Intr	73554	73513	-	0
20	Intr	73679	73638	-	0
	Intr	73857	73758	-	0
	Intr	74009	73942	-	0
	Intr	74203	74119	-	0
	Init	75204	74990	-	0
25	>2760164	/122489			
	len =	748	nex =	3	
30	Term	1929	1667	-	0
	Intr	2065	2000	-	0
	Init	2414	2235	-	0
	>2760164	/141953			
35	len =	404	nex =	1	
	Sngl	40494	40091	-	0
40	>2760164	/6393			
	len =	1887	nex =	4	
	Term	45207	44626	-	0
45	Intr	45909	45643	-	0
	Intr	46332	46110	-	0
	Init	46512	46432	-	0
	>2760164	/12030			
50	len =	1175	nex =	1	
	Sngl	62893	63617	+	0
55	>2760164	/5684			
	len =	1901	nex =	6	
	Init	71453	71775	+	0
60	Intr	71875	71990	+	0

1020

	Intr	72555	72678	+	0
	Intr	72782	72930	+	0
	Intr	73031	73093	+	0
5	Term	73185	73353	+	0
	>2760164 /36046				
	len =	910	nex =	3	
10	Init	871	1095	+	0
	Intr	1189	1297	+	0
	Term	1413	1767	+	0
	>2760164 /6826				
15	len =	758	nex =	3	
	Init	991	1095	+	0
	Intr	1189	1297	+	0
20	Term	1413	1748	+	0
	>2760165 /19114				
25	len =	1646	nex =	1	
	Sngl	20275	19212	-	0
	>2760165 /4193				
30	len =	1620	nex =	5	
	Init	30702	30854	+	0
	Intr	30939	31019	+	0
	Intr	31132	31326	+	0
35	Intr	31420	31797	+	0
	Term	31892	32219	+	0
	>2760165 /17486				
40	len =	386	nex =	1	
	Sngl	32921	32536	-	0
	>2760165 /17455				
45	len =	1330	nex =	4	
	Term	32949	32576	-	0
	Intr	33144	33034	-	0
50	Intr	33293	33222	-	0
	Init	33897	33392	-	0
	>2760165 /2869				
55	len =	2553	nex =	6	
	Term	32949	32636	-	0
	Intr	33144	33034	-	0
	Intr	33293	33222	-	0
60	Intr	34521	33392	-	0

2025 RELEASE UNDER E.O. 14176

1021

	Intr	34888	34819	-	0
	Init	35188	34989	-	0
5 >2760165 /1022					
	len =	1103	nex =	3	
	Term	37453	37053	-	0
10	Intr	37627	37537	-	0
	Init	38155	37697	-	0
15 >2760165 /35974					
	len =	1669	nex =	4	
	Init	42735	42859	+	0
	Intr	43198	43462	+	0
	Intr	43554	43748	+	0
20	Term	43823	44403	+	0
25 >2760165 /34861					
	len =	2759	nex =	4	
	Term	45311	44615	-	0
	Intr	45749	45508	-	0
	Intr	46640	46252	-	0
	Init	47373	47034	-	0
30	>2760165 /146274				
	len =	1378	nex =	3	
	Init	56343	56430	+	0
35	Intr	56568	56695	+	0
	Term	56823	57131	+	0
40 >2760165 /2996					
	len =	1590	nex =	3	
	Init	6987	7404	+	0
	Intr	7510	8043	+	0
45	Term	8152	8576	+	0
50 >2760166 /37308					
	len =	1640	nex =	3	
	Init	36323	36769	+	0
	Intr	36888	37165	+	0
	Term	37289	37962	+	0
55 >2760166 /37792					
	len =	1517	nex =	2	
	Init	36888	37165	+	0
60	Term	37289	37987	+	0

>2760166 /15419

len = 762 nex = 2

5	Term	5587	5255	-	0
	Init	6016	5725	-	0

>2760167 /6274

10 len = 2810 nex = 11

	Term	32480	31948	-	0
	Intr	32689	32573	-	0
	Intr	32895	32777	-	0
15	Intr	33101	32993	-	0
	Intr	33238	33200	-	0
	Intr	33451	33373	-	0
	Intr	33637	33563	-	0
	Intr	33811	33747	-	0
20	Intr	33992	33937	-	0
	Intr	34225	34133	-	0
	Init	34757	34493	-	0

>2760167 /101988

len = 1033 nex = 1

Sngl 40367 40745 + 0

30 >2760167 /92179

len = 993 nex = 1

Sngl 40367 40745 + 0

35 >2760167 /13697

len = 1356 nex = 3

40	Term	42553	42287	-	0
	Intr	42795	42646	-	0
	Init	43642	43418	-	0

>2760167 /30674

len = 804 nex = 3

	Term	44295	44028	-	0
	Intr	44536	44458	-	0
50	Init	44831	44714	-	0

>2760167 /7119

len = 1930 nex = 3

55	Init	47503	47617	+	0
	Intr	47702	48982	+	0
	Term	49233	49428	+	0

60 >2760167 /9480

2025-09-26 10:00:00

1023

	len =	1375	nex =	6	
	Init	51156	51210	+	0
5	Intr	51501	51557	+	0
	Intr	51659	51776	+	0
	Intr	51994	52083	+	0
	Term	52170	52349	+	0
	Term	52439	52514	+	0
10	>2760167 /35248				
	len =	116	nex =	1	
15	Sngl	52672	52787	+	0
	>2760167 /20790				
	len =	1918	nex =	6	
20	Init	53351	53565	+	0
	Intr	53877	54048	+	0
	Intr	54137	54333	+	0
	Intr	54434	54601	+	0
25	Intr	54690	54818	+	0
	Term	54954	55268	+	0
	>2760167 /8984				
30	len =	3272	nex =	14	
	Init	6342	6410	+	0
	Intr	6499	6570	+	0
	Intr	6719	6773	+	0
35	Intr	6858	6876	+	0
	Intr	6960	7027	+	0
	Intr	7214	7288	+	0
	Intr	7382	7437	+	0
	Intr	7542	7586	+	0
40	Intr	7667	7930	+	0
	Intr	8046	8140	+	0
	Intr	8260	8350	+	0
	Intr	8439	8549	+	0
	Intr	8647	8841	+	0
45	Term	8935	9304	+	0
	>2760167 /37493				
	len =	3313	nex =	4	
50	Term	66659	66266	-	0
	Intr	68318	68043	-	0
	Intr	69011	68592	-	0
	Init	69578	69461	-	0
55	>2760167 /7548				
	len =	1930	nex =	6	
60	Term	79732	79374	-	0

1024

	Intr	80062	79829	-	0
	Intr	80336	80169	-	0
	Intr	80655	80426	-	0
	Intr	80848	80749	-	0
5	Init	81298	80951	-	0
>2760167 /10973					
10	len =	1259	nex =	1	
	Sngl	82221	82174	-	0
>2760167 /10749					
15	len =	1330	nex =	3	
	Term	81820	81596	-	0
	Intr	82221	82174	-	0
	Init	82918	82706	-	0
20	>2760168 /1165				
	len =	1197	nex =	3	
25	Init	13720	13876	+	0
	Intr	14260	14352	+	0
	Term	14686	14916	+	0
>2760168 /42577					
30	len =	796	nex =	2	
	Term	23611	23112	-	0
	Init	23907	23692	-	0
35	>2760168 /12729				
	len =	216	nex =	1	
40	Sngl	2967	2752	-	0
>2760168 /19343					
45	len =	612	nex =	1	
	Sngl	2989	2378	-	0
>2760168 /17242					
50	len =	2173	nex =	7	
	Init	31582	31714	+	0
	Intr	31798	31982	+	0
	Intr	32205	32264	+	0
55	Intr	32354	32635	+	0
	Intr	32868	32963	+	0
	Intr	33072	33268	+	0
	Term	33388	33754	+	0
60	>2760168 /94968				

20250901 09:49:03

1025

	len =	467	nex =	1	
5	Sngl	3825	4291	+	0
	>2760168	/152076			
	len =	1210	nex =	3	
10	Init	598	955	+	0
	Intr	1255	1356	+	0
	Term	1441	1806	+	0
	>2760169	/8993			
15	len =	2713	nex =	11	
	Init	1	127	+	0
	Intr	474	621	+	0
20	Intr	698	769	+	0
	Intr	943	1014	+	0
	Intr	1227	1298	+	0
	Intr	1396	1539	+	0
	Intr	1636	1716	+	0
25	Intr	1844	1886	+	0
	Intr	1966	2037	+	0
	Intr	2143	2214	+	0
	Term	2426	2713	+	0
30	>2760169	/14492			
	len =	651	nex =	2	
	Init	20745	20918	+	0
35	Term	21014	21373	+	0
	>2760169	/5810			
	len =	1439	nex =	1	
40	Sngl	24319	23980	-	0
	>2760169	/38421			
45	len =	1851	nex =	1	
	Sngl	44204	42354	-	0
	>2760169	/16827			
50	len =	564	nex =	1	
	Sngl	48497	49060	+	0
55	>2760169	/12970			
	len =	1536	nex =	3	
	Term	72431	71753	-	0
60	Intr	72794	72517	-	0

2025-06-16 09:22:52

1026

	Init	73288	72925	-	0
	>2760170	/36309			
5	len =	2470	nex =	7	
	Term	8439	7872	-	0
	Intr	8753	8655	-	0
	Intr	8923	8846	-	0
10	Intr	9409	9327	-	0
	Intr	9622	9525	-	0
	Intr	9990	9959	-	0
	Init	10341	10081	-	0
15	>2760170	/100904			
	len =	616	nex =	2	
	Term	72185	72007	-	0
20	Init	72622	72442	-	0
	>2760170	/4536			
	len =	706	nex =	2	
25	Term	72185	71972	-	0
	Init	72677	72442	-	0
	>2760170	/292			
30	len =	684	nex =	2	
	Term	72185	72004	-	0
	Init	72687	72442	-	0
35	>2760170	/99615			
	len =	1994	nex =	5	
40	Init	79105	79274	+	0
	Intr	79891	79988	+	0
	Intr	80281	80424	+	0
	Intr	80520	80749	+	0
	Term	80836	81098	+	0
45	>2760171	/4585			
	len =	2746	nex =	12	
50	Term	25736	25556	-	0
	Intr	25904	25816	-	0
	Intr	26088	25978	-	0
	Intr	26440	26311	-	0
	Intr	26630	26532	-	0
55	Intr	26871	26767	-	0
	Intr	27087	26953	-	0
	Intr	27378	27174	-	0
	Intr	27618	27490	-	0
	Intr	27801	27711	-	0
60	Intr	27896	27863	-	0

1027

	Init	28301	28131	-	0
	>2760171		/12301		
5	len =	1349	nex =	4	
	Init	51419	51657	+	0
	Intr	51763	51863	+	0
	Intr	51971	52167	+	0
10	Term	52288	52767	+	0
	>2760171		/31563		
15	len =	2637	nex =	4	
	Init	7482	7708	+	0
	Intr	7821	7962	+	0
	Intr	8116	8177	+	0
20	Term	8473	8676	+	0
	>2760171		/37100		
	len =	1735	nex =	8	
25	Init	69229	69386	+	0
	Intr	69531	69707	+	0
	Intr	69794	69925	+	0
	Intr	70009	70107	+	0
	Intr	70188	70304	+	0
30	Intr	70399	70494	+	0
	Intr	70648	70760	+	0
	Term	70850	70954	+	0
	>2760172		/24599		
35	len =	730	nex =	4	
	Term	26279	26046	-	0
	Intr	26489	26370	-	0
40	Intr	26593	26564	-	0
	Init	26767	26693	-	0
	>2760172		/25884		
45	len =	670	nex =	1	
	Sngl	51002	51664	+	0
	>2760172		/8981		
50	len =	357	nex =	1	
	Sngl	51004	51353	+	0
55	>2760172		/21903		
	len =	940	nex =	1	
60	Sngl	57422	58361	+	0

20250910 14:28:59

1028

>2760172 /14033

len = 1585 nex = 4

5	Term	71969	71571	-	0
	Intr	72172	72083	-	0
	Intr	72354	72271	-	0
	Init	73155	72644	-	0

10 >2760172 /205721

len = 693 nex = 1

Sngl 78960 78268 - 0

15 >2760173 /41408

len = 1404 nex = 3

20	Init	2606	2664	+	0
	Intr	2991	3372	+	0
	Term	3572	4009	+	0

25 >2760173 /100590

len = 1692 nex = 5

	Init	27003	27364	+	0
	Intr	27452	27570	+	0
30	Intr	27661	27790	+	0
	Intr	27912	28130	+	0
	Term	28473	28694	+	0

35 >2760173 /14898

len = 1241 nex = 3

	Term	43600	43275	-	0
	Intr	44232	44110	-	0
40	Init	44515	44319	-	0

>2760173 /7791

len = 1186 nex = 3

45	Term	43600	43330	-	0
	Intr	44232	44110	-	0
	Init	44515	44319	-	0

50 >2760173 /15331

len = 1000 nex = 3

	Term	43600	43516	-	0
55	Intr	44232	44110	-	0
	Init	44515	44319	-	0

>2760173 /4734

60 len = 1047 nex = 2

1030

	>2760829	/33692		
5	len =	1822	nex =	3
	Init	11711	12231	+
	Intr	12887	13110	+
	Term	13186	13532	+
10	>2760829	/41755		
	len =	1546	nex =	2
	Term	14354	13832	-
15	Init	15377	14681	-
	>2760829	/28786		
20	len =	1722	nex =	1
	Sngl	70	1791	+
	>2760829	/4754		
25	len =	970	nex =	2
	Term	77740	77286	-
	Init	77876	77815	-
30	>2795802	/17353		
	len =	1734	nex =	3
	Term	20024	19317	-
35	Intr	20663	20345	-
	Init	21050	20759	-
	>2795802	/27620		
40	len =	670	nex =	1
	Sngl	32381	33046	+
45	>2815404	/42968		
	len =	610	nex =	1
	Sngl	14332	14937	+
50	>2815404	/22971		
	len =	620	nex =	1
	Sngl	14341	14960	+
55	>2815404	/9731		
	len =	675	nex =	1
60	Sngl	35713	35039	-

2025 RELEASE UNDER E.O. 14176

	>2815404	/18451		
5	len =	2372	nex =	8
	Init	45075	45423	+
	Intr	45497	45617	+
	Intr	45704	46131	+
	Intr	46227	46383	+
10	Intr	46695	46777	+
	Intr	46860	46956	+
	Intr	47066	47135	+
	Term	47312	47446	+
15	>2815404	/15499		
	len =	1870	nex =	4
	Init	67484	67838	+
20	Intr	68518	68746	+
	Intr	68868	68992	+
	Term	69083	69344	+
25	>2815404	/18800		
	len =	910	nex =	2
	Init	70956	71114	+
	Term	71453	71857	+
30	>2815404	/9878		
	len =	687	nex =	1
35	Sngl	77572	76886	-
	>2815519	/96891		
	len =	1073	nex =	6
40	Term	22493	22461	-
	Intr	22634	22576	-
	Intr	22953	22881	-
	Intr	23103	23039	-
45	Intr	23313	23201	-
	Init	23533	23402	-
	>2815519	/25380		
50	len =	750	nex =	1
	Sngl	26997	26248	-
	>2815519	/8432		
55	len =	310	nex =	1
	Sngl	27005	26703	-
60	>2815519	/36424		

1032

	len =	412	nex =	1	
5	Sngl	27008	26597	-	0
	>2815519	/2327			
	len =	394	nex =	1	
10	Sngl	27008	26615	-	0
	>2815519	/113981			
	len =	819	nex =	1	
15	Sngl	27010	26192	-	0
	>2815519	/7630			
20	len =	763	nex =	1	
	Sngl	27010	26248	-	0
	>2815519	/31731			
25	len =	762	nex =	1	
	Sngl	27010	26249	-	0
30	>2815519	/35047			
	len =	396	nex =	1	
	Sngl	27010	26615	-	0
35	>2815519	/16438			
	len =	400	nex =	1	
40	Sngl	27014	26615	-	0
	>2815519	/36235			
	len =	773	nex =	1	
45	Sngl	27020	26248	-	0
	>2815519	/20963			
50	len =	2032	nex =	9	
	Init	44481	44658	+	0
	Intr	44737	44883	+	0
	Intr	44989	45105	+	0
55	Intr	45281	45424	+	0
	Intr	45505	45591	+	0
	Intr	45673	45788	+	0
	Intr	45891	46072	+	0
	Intr	46172	46229	+	0
60	Term	46318	46512	+	0

2025 RELEASE UNDER E.O. 14176

1033

>2815519 /4249

```

5      len =      1602      nex =      5
      Init  47167  47260      +      0
      Intr  47353  47634      +      0
      Intr  47724  47945      +      0
      Intr  48032  48196      +      0
10     Term  48281  48768      +      0

```

>2815519 /11560

15	len =	1777	nex =	4	
	Init	61717	61829	+	0
	Intr	61954	62030	+	0
	Intr	62992	63183	+	0
	Term	63276	63493	+	0

>2815519 /113946

```

len =      504    nex =      2
25      Init  62991  63183      +      0
      Term  63276  63494      +      0

```

>2815519 /36710

```

30      len =      414      nex =          1
          Snql 76124      75711      -          0

```

>2815519 /14108

```
len =      1046    nex =          1
Snq1 76756    75711    ~          0
```

40 >2815519 /38663

```
len =      1041      nex =          1
      Sngl 76756    75716      -          0
```

>2815519 /11318

```

len =      1438    nex =      2
50  Term  76754    75679    -      0
    Init  77108    77052    -      0

```

>2815519 /38261

```

55      len =      1484      nex =          2
      Term 76754  75635      -      0
      Init 77108  77052      -      0

```

60 >2815519 /25772

1034

	len =	1427	nex =	1	
5	Sngl	77108	77052	-	0
	>2815519	/6957			
	len =	2110	nex =	5	
10	Init	89813	89995	+	0
	Intr	90317	90488	+	0
	Intr	90587	91007	+	0
	Intr	91091	91156	+	0
	Term	91508	91915	+	0
15	>2815519	/32361			
	len =	2950	nex =	12	
20	Term	92911	92505	-	0
	Intr	93238	92998	-	0
	Intr	93386	93324	-	0
	Intr	93536	93468	-	0
	Intr	93725	93642	-	0
25	Intr	94016	93811	-	0
	Intr	94196	94106	-	0
	Intr	94389	94285	-	0
	Intr	94796	94722	-	0
	Intr	94970	94878	-	0
30	Intr	95139	95064	-	0
	Init	95447	95254	-	0
	>2827513	/34120			
35	len =	2154	nex =	3	
	Term	16992	16242	-	0
	Intr	17554	17075	-	0
	Init	18395	18232	-	0
40	>2827513	/24334			
	len =	1775	nex =	3	
45	Term	23483	22952	-	0
	Intr	24280	23812	-	0
	Init	24726	24647	-	0
	>2827513	/39740			
50	len =	2470	nex =	7	
	Init	26839	26915	+	0
	Intr	27027	27083	+	0
55	Intr	27164	27220	+	0
	Intr	27310	27350	+	0
	Intr	27467	27538	+	0
	Intr	27635	27714	+	0
	Term	27850	28215	+	0
60					

>2827513 /24184

len = 190 nex = 1

5 Sngl 28607 28793 + 0

>2827513 /229

len = 2269 nex = 8

10 Init 28607 28791 + 0

Intr 28967 29043 + 0

Intr 29236 29291 + 0

Intr 29386 29458 + 0

15 Intr 29960 30071 + 0

Intr 30213 30286 + 0

Intr 30393 30467 + 0

Term 30566 30875 + 0

20 >2827513 /39364

len = 402 nex = 1

25 Sngl 42615 43016 + 0

>2827513 /27911

len = 4030 nex = 12

30 Init 45284 45583 + 0

Intr 45967 46094 + 0

Intr 46215 46286 + 0

Intr 46380 46493 + 0

35 Intr 46958 47026 + 0

Intr 47094 47150 + 0

Intr 47288 47341 + 0

Intr 47427 47489 + 0

Intr 47586 47648 + 0

40 Intr 47783 47938 + 0

Intr 48041 48196 + 0

Term 48315 48388 + 0

>2827513 /19281

45 len = 1222 nex = 2

Init 48103 48196 + 0

Term 48315 48388 + 0

50 >2827513 /3172

len = 414 nex = 1

55 Sngl 69669 70082 + 0

>2827513 /13678

len = 696 nex = 0

60 >2827513 /9280

1036

len = 236 nex = 1

5 Sngl 82731 82496 - 0

>2827538 /107860

len = 756 nex = 2

10 Init 10127 10535 + 0

Term 10626 10882 + 0

>2827538 /35786

15 len = 2746 nex = 10

Init 11131 11383 + 0

Intr 11755 11809 + 0

Intr 11898 12053 + 0

20 Intr 12143 12311 + 0

Intr 12391 12455 + 0

Intr 12534 12674 + 0

Intr 12769 12885 + 0

Intr 13024 13380 + 0

25 Intr 13468 13525 + 0

Term 13612 13876 + 0

>2827538 /22353

30 len = 1781 nex = 1

Sngl 22916 24696 + 0

>2827538 /41062

35 len = 1775 nex = 3

Init 30074 30181 + 0

Intr 31148 31303 + 0

40 Term 31408 31848 + 0

>2827538 /20948

45 len = 2012 nex = 6

Term 37215 36924 - 0

Intr 37385 37302 - 0

Intr 37555 37480 - 0

Intr 37720 37640 - 0

50 Intr 38064 37811 - 0

Init 38618 38156 - 0

>2827538 /38360

55 len = 2206 nex = 7

Term 42538 42224 - 0

Intr 42735 42673 - 0

Intr 42925 42840 - 0

60 Intr 43152 43023 - 0

1037

Intr	43349	43258	-	0
Intr	43650	43437	-	0
Init	44429	44185	-	0

5 >2827538 /39135

len = 1970 nex = 5

Term	48221	47882	-	0
Intr	48414	48382	-	0
Intr	48655	48508	-	0
Intr	49319	48997	-	0
Init	49851	49398	-	0

15 >2827538 /12041

len = 1958 nex = 4

Term	57306	56734	-	0
Intr	57614	57483	-	0
Intr	57766	57699	-	0
Init	58691	58063	-	0

25 >2827538 /13832

len = 1234 nex = 2

Term	81460	80806	-	0
Init	82039	81838	-	0

30 >2827538 /9746

len = 2173 nex = 7

Term	84745	84562	-	0
Intr	84904	84822	-	0
Intr	85251	84998	-	0
Intr	85601	85349	-	0
Intr	85852	85679	-	0
Intr	86443	86358	-	0
Init	86734	86552	-	0

>2827538 /18053

45 len = 1301 nex = 1

Sngl 9200 7900 - 0

>2827644 /40232

50 len = 2013 nex = 7

Init	50573	51055	+	0
Intr	51150	51288	+	0
Intr	51431	51539	+	0
Intr	51732	51841	+	0
Intr	51956	52103	+	0
Intr	52222	52286	+	0
Term	52373	52585	+	0

60

1038

>2827644 /15834

```
len = 2350    nex = 3
```

5	Term	59427	58305	-	0
	Intr	60290	60054	-	0
	Init	60647	60399	-	0

>2827644 /32637

```

10      len =      2027      nex =      9

```

	Init	88415	88513	+	0
15	Intr	88598	88726	+	0
	Intr	88827	88916	+	0
	Intr	89005	89155	+	0
	Intr	89262	89371	+	0
20	Intr	89447	89518	+	0
	Intr	89612	89717	+	0
	Intr	89802	89996	+	0
	Term	90088	90441	+	0

>2827698 /29775

```
25      len =      2250      nex =          6
```

	Init	21347	21716	+	0
	Intr	21801	21847	+	0
	Intr	21943	22057	+	0
30	Intr	22752	22796	+	0
	Intr	23036	23131	+	0
	Term	23213	23596	+	0

>2827698 /17160

```

35      len =      1210      nex =      1

```

Sng1	24108	25313	+	0
------	-------	-------	---	---

40 >2827698 /36946

```
len = 2069 nex = 5
```

45	Init	28482	28521	+	0
	Intr	28627	28665	+	0
	Intr	28798	29172	+	0
	Intr	29262	29317	+	0
	Term	29444	29760	+	0

50 >2827698 /16111

```
len = 1796    nex = 3
```

	Init	3992	4615	+	0
55	Intr	4773	4960	+	0
	Term	5255	5787	+	0

>2828180 /31285

```
60      len =      1117      nex =          3
```


1039

	Init	41115	41311	+	0
	Intr	41410	41571	+	0
	Term	41636	42231	+	0
5	>2828180 /9396				
	len =	500	nex =	1	
10	Sngl	41747	42246	+	0
	>2828180 /115921				
	len =	392	nex =	1	
15	Sngl	41841	42232	+	0
	>2828180 /32771				
20	len =	2958	nex =	12	
	Init	53810	53931	+	0
	Intr	54276	54407	+	0
	Intr	54599	54657	+	0
25	Intr	54939	55069	+	0
	Intr	55154	55239	+	0
	Intr	55348	55379	+	0
	Intr	55608	55769	+	0
	Intr	55869	55941	+	0
30	Intr	56027	56127	+	0
	Intr	56276	56368	+	0
	Intr	56471	56517	+	0
	Term	56610	56767	+	0
35	>2828182 /5344				
	len =	1193	nex =	2	
	Init	11326	11889	+	0
40	Term	11959	12518	+	0
	>2828182 /35527				
	len =	1242	nex =	2	
45	Init	12894	13457	+	0
	Term	13541	14135	+	0
	>2828182 /207083				
50	len =	1835	nex =	3	
	Term	60046	59494	-	0
	Intr	60847	60710	-	0
55	Init	61328	61093	-	0
	>2828182 /17482				
60	len =	1854	nex =	3	

1040

Term	60046	59509	-	0
Intr	60847	60710	-	0
Init	61362	61093	-	0

5 >2828182 /158804

len = 501 nex = 1

Sngl 62845 62345 - 0

10 >2828182 /15560

len = 2329 nex = 3

Term	62937	62347	-	0
Intr	63667	63187	-	0
Init	64675	64169	-	0

>2828182 /17411

20 len = 1728 nex = 2

Init	79315	79968	+	0
Term	80239	81042	+	0

25 >2828182 /13813

len = 1618 nex = 4

Init	81708	81785	+	0
Intr	81894	82074	+	0
Intr	82162	82259	+	0
Term	82353	82976	+	0

35 >2828182 /22460

len = 850 nex = 1

Sngl 85041 85888 + 0

40 >2828183 /9655

len = 1611 nex = 8

Init	15617	15682	+	0
Intr	15785	15864	+	0
Intr	15960	16044	+	0
Intr	16138	16185	+	0
Intr	16274	16342	+	0
Intr	16428	16583	+	0
Intr	16725	16818	+	0
Term	16920	17227	+	0

>2828183 /15495

55 len = 610 nex = 1

Sngl 2072 1467 - 0

60 >2828183 /38900

20250707 09:25:52

1041

	len =	2564	nex =	7	
	Term	33533	33168	-	0
5	Intr	33682	33616	-	0
	Intr	33903	33802	-	0
	Intr	34116	34003	-	0
	Intr	34399	34226	-	0
	Intr	35100	34669	-	0
10	Init	35731	35192	-	0
>2828183		/18973			
	len =	405	nex =	1	
15	Sngl	39229	38825	-	0
>2828184		/118036			
20	len =	658	nex =	2	
	Init	14494	14634	+	0
	Term	14779	15151	+	0
25	>2828184	/11922			
	len =	1978	nex =	6	
	Init	42361	42487	+	0
30	Intr	42866	42902	+	0
	Intr	43191	43413	+	0
	Intr	43509	43670	+	0
	Intr	43775	43897	+	0
	Term	43999	44338	+	0
35	>2828184	/17188			
	len =	3468	nex =	12	
40	Init	54034	54226	+	0
	Intr	54348	54434	+	0
	Intr	54570	54675	+	0
	Intr	54848	54883	+	0
	Intr	55623	55674	+	0
45	Intr	55958	56022	+	0
	Intr	56150	56206	+	0
	Intr	56372	56444	+	0
	Intr	56529	56615	+	0
	Intr	56712	56877	+	0
50	Intr	56959	57068	+	0
	Term	57147	57481	+	0
>2828185		/102248			
55	len =	1092	nex =	2	
	Init	13901	14405	+	0
	Term	14538	14992	+	0
60	>2828185	/14679			

2025-09-16 09:16:00

1042

	len =	501	nex =	1	
5	Sngl	19843	20343	+	0
	>2828185	/33972			
	len =	2385	nex =	9	
10	Init	53063	53348	+	0
	Intr	53686	53760	+	0
	Intr	53854	53955	+	0
	Intr	54029	54082	+	0
	Intr	54183	54275	+	0
15	Intr	54389	54481	+	0
	Intr	54563	54667	+	0
	Intr	54744	54842	+	0
	Term	54920	55447	+	0
20	>2828185	/3988			
	len =	1995	nex =	9	
25	Init	53214	53348	+	0
	Intr	53686	53760	+	0
	Intr	53854	53955	+	0
	Intr	54029	54082	+	0
	Intr	54183	54275	+	0
	Intr	54389	54481	+	0
30	Intr	54536	54667	+	0
	Intr	54744	54842	+	0
	Term	54920	55208	+	0
	>2828185	/15373			
35	len =	1398	nex =	6	
	Init	68405	68507	+	0
	Intr	68622	68658	+	0
40	Intr	68914	68999	+	0
	Intr	69103	69198	+	0
	Intr	69384	69546	+	0
	Term	69646	69802	+	0
45	>2828186	/108599			
	len =	1481	nex =	4	
	Init	14950	15033	+	0
50	Intr	15143	15301	+	0
	Intr	15523	15749	+	0
	Term	15975	16194	+	0
	>2828186	/20057			
55	len =	1042	nex =	2	
	Term	42331	42168	-	0
	Init	43209	43006	-	0
60					

2025 RELEASE UNDER E.O. 14176

1043

>2828186 /116465

len = 884 nex = 1

5 Sngl 43436 44319 + 0

>2828186 /36363

len = 1150 nex = 1

10 Sngl 53511 53675 + 0

>2828186 /117347

15 len = 791 nex = 1

Sngl 60600 61390 + 0

>2828186 /6461

20 len = 859 nex = 3

Init 71764 72268 + 0

Intr 72381 72437 + 0

25 Term 72534 72612 + 0

>2828187 /23099

30 len = 1114 nex = 2

Term 64086 63766 - 0

Init 64879 64636 - 0

>2828187 /9905

35 len = 1351 nex = 6

Term 72205 71968 - 0

Intr 72362 72299 - 0

40 Intr 72523 72492 - 0

Intr 72708 72611 - 0

Intr 73206 73071 - 0

Init 73318 73233 - 0

45 >2828187 /91844

len = 1425 nex = 5

Term 72205 71961 - 0

50 Intr 72362 72299 - 0

Intr 72523 72492 - 0

Intr 72708 72611 - 0

Init 73206 73071 - 0

55 >2828187 /142899

len = 1418 nex = 5

Term 72205 71968 - 0

60 Intr 72362 72299 - 0

2025 RELEASE UNDER E.O. 14176

1044

Intr	72523	72492	-	0
Intr	72708	72611	-	0
Init	73206	73071	-	0

5	>2828188	/17485		
	len =	879	nex =	3
10	Init	27707	27799	+
	Intr	27901	28043	+
	Term	28241	28522	+
	>2828188	/1983		
15	len =	2312	nex =	10
	Init	30424	30603	+
	Intr	30707	30804	+
	Intr	30896	30993	+
20	Intr	31098	31147	+
	Intr	31561	31656	+
	Intr	31753	31842	+
	Intr	31914	32006	+
	Intr	32089	32184	+
25	Intr	32288	32353	+
	Term	32425	32735	+
	>2828188	/2949		
30	len =	506	nex =	1
	Sngl	37555	37050	-
	>2828188	/34409		
35	len =	1630	nex =	2
	Init	49022	49121	+
	Term	49382	50625	+
40	>2828188	/101070		
	len =	672	nex =	1
45	Sngl	68221	68889	+
	>2828278	/93971		
	len =	630	nex =	1
50	Sngl	20720	20517	-
	>2828278	/9641		
55	len =	1690	nex =	2
	Term	25562	25106	-
	Init	26162	26038	-
60	>2828278	/3981		

20250910-091600

1045

	len =	1341	nex =	3	
	Term	28518	27857	-	0
5	Intr	28679	28602	-	0
	Init	29197	28762	-	0
	>2828278 /101094				
10	len =	261	nex =	1	
	Sngl	47484	47744	+	0
	>2828278 /36576				
15	len =	2509	nex =	7	
	Term	57030	56730	-	0
	Intr	57325	57174	-	0
20	Intr	57571	57416	-	0
	Intr	57973	57688	-	0
	Intr	58172	58056	-	0
	Intr	58336	58274	-	0
	Init	59238	58896	-	0
25	>2832611 /6321				
	len =	2846	nex =	2	
30	Init	12902	13264	+	0
	Term	14896	15747	+	0
	>2832611 /115884				
35	len =	2174	nex =	5	
	Init	26411	26644	+	0
	Intr	26778	27002	+	0
	Intr	27714	27932	+	0
40	Intr	28070	28131	+	0
	Term	28235	28584	+	0
	>2832611 /8768				
45	len =	1456	nex =	4	
	Init	29030	29347	+	0
	Intr	29729	29811	+	0
	Intr	29893	30007	+	0
50	Term	30105	30485	+	0
	>2832611 /38712				
55	len =	2988	nex =	12	
	Init	35656	36292	+	0
	Intr	36401	36475	+	0
	Intr	36567	36662	+	0
	Intr	36750	36835	+	0
60	Intr	36921	37041	+	0

1046

	Intr	37124	37203	+	0
	Intr	37298	37360	+	0
	Intr	37467	37573	+	0
	Intr	37704	37789	+	0
5	Intr	37867	37988	+	0
	Intr	38095	38206	+	0
	Term	38306	38643	+	0
	>2832611 /92096				
10	len =	1527	nex =	6	
	Term	67967	67742	-	0
	Intr	68134	68059	-	0
15	Intr	68294	68206	-	0
	Intr	68482	68429	-	0
	Intr	68719	68580	-	0
	Init	68873	68799	-	0
20	>2832611 /8996				
	len =	1431	nex =	4	
	Term	70354	70256	-	0
25	Intr	70528	70448	-	0
	Intr	71169	70723	-	0
	Init	71686	71473	-	0
	>2832611 /109513				
30	len =	1954	nex =	4	
	Term	82219	81803	-	0
	Intr	82755	82537	-	0
35	Intr	83404	83180	-	0
	Init	83756	83517	-	0
	>2832639 /155370				
40	len =	809	nex =	4	
	Init	13801	13917	+	0
	Intr	13999	14045	+	0
	Intr	14121	14217	+	0
45	Term	14310	14609	+	0
	>2832639 /16226				
	len =	1821	nex =	0	
50	>2832639 /10320				
	len =	580	nex =	1	
55	Sngl	22887	22331	-	0
	>2832639 /24883				
60	len =	490	nex =	1	

1047

	Sngl	28760	29244	+	0
	>2832639 /14794				
5	len =	1378	nex =	3	
	Term	29423	29196	-	0
	Intr	29908	29796	-	0
	Init	30573	30411	-	0
10	>2832639 /907				
	len =	802	nex =	1	
15	Sngl	41456	40655	-	0
	>2832639 /10310				
20	len =	2754	nex =	15	
	Init	52225	52432	+	0
	Intr	52507	52573	+	0
	Intr	52668	52761	+	0
	Intr	52853	52909	+	0
25	Intr	53005	53074	+	0
	Intr	53159	53223	+	0
	Intr	53319	53400	+	0
	Intr	53494	53543	+	0
	Intr	53622	53765	+	0
30	Intr	53843	53913	+	0
	Intr	53979	54048	+	0
	Intr	54133	54198	+	0
	Intr	54315	54473	+	0
	Intr	54555	54652	+	0
35	Term	54755	54978	+	0
	>2832639 /15817				
40	len =	753	nex =	1	
	Sngl	61848	61096	-	0
	>2832639 /7501				
45	len =	1667	nex =	3	
	Term	62268	61093	-	0
	Intr	62493	62378	-	0
	Init	62759	62575	-	0
50	>2832639 /35285				
	len =	850	nex =	4	
55	Init	8531	8838	+	0
	Intr	8913	9050	+	0
	Intr	9136	9198	+	0
	Term	9284	9373	+	0
60	>2832639 /31734				

1048

	len =	1358	nex =	4	
5	Term	94255	94004	-	0
	Intr	94442	94354	-	0
	Intr	94664	94560	-	0
	Init	94843	94774	-	0
	>2832639 /24295				
10	len =	926	nex =	2	
	Init	95837	96095	+	0
	Term	96192	96466	+	0
15	>2832639 /29009				
	len =	1136	nex =	1	
20	Sngl	97830	98023	+	0
	>2832667 /113536				
25	len =	701	nex =	2	
	Init	14746	14852	+	0
	Term	14974	15446	+	0
	>2832667 /26401				
30	len =	881	nex =	2	
	Init	16072	16194	+	0
	Term	16286	16952	+	0
35	>2832667 /8156				
	len =	1771	nex =	3	
40	Term	20352	19919	-	0
	Intr	20953	20899	-	0
	Init	21689	21342	-	0
	>2832667 /19552				
45	len =	1788	nex =	3	
	Term	20352	19913	-	0
	Intr	20953	20899	-	0
50	Init	21700	21342	-	0
	>2832667 /4390				
	len =	2022	nex =	2	
55	Init	24444	24552	+	0
	Term	24642	25004	+	0
	>2832667 /20563				
60					

1049

	len =	337	nex =	1	
	Sngl	24653	24317	-	0
5	>2832667	/33086			
	len =	578	nex =	1	
10	Sngl	27162	26726	-	0
	>2832667	/111376			
	len =	1337	nex =	3	
15	Term	27162	26726	-	0
	Intr	27443	27248	-	0
	Init	28062	27564	-	0
	>2832667	/41007			
20	len =	1362	nex =	3	
	Term	27162	26726	-	0
	Intr	27443	27248	-	0
25	Init	28083	27564	-	0
	>2832667	/6848			
30	len =	1157	nex =	1	
	Sngl	3909	2753	-	0
	>2832667	/5132			
35	len =	1474	nex =	5	
	Init	52118	52255	+	0
	Intr	52656	52754	+	0
	Intr	52930	52982	+	0
40	Intr	53135	53208	+	0
	Term	53464	53591	+	0
	>2832667	/35323			
45	len =	1474	nex =	5	
	Init	52120	52255	+	0
	Intr	52656	52754	+	0
	Intr	52930	52982	+	0
50	Intr	53135	53208	+	0
	Term	53464	53593	+	0
	>2832667	/4495			
55	len =	2114	nex =	1	
	Sngl	59794	60466	+	0
60	>2832667	/2081			

1050

	len =	1232	nex =	3	
	Init	68491	69116	+	0
	Intr	69422	69533	+	0
5	Term	69623	69722	+	0
	>2832667 /12272				
	len =	2316	nex =	5	
10	Init	68540	69116	+	0
	Intr	69422	69533	+	0
	Intr	69623	69794	+	0
	Intr	69883	70085	+	0
15	Term	70296	70411	+	0
	>2832667 /2673				
	len =	320	nex =	1	
20	Sngl	72959	72640	-	0
	>2832689 /2075				
	len =	595	nex =	1	
25	Sngl	42824	43418	+	0
	>2832689 /5753				
	len =	730	nex =	1	
30	Sngl	42881	43602	+	0
	>2832689 /33965				
	len =	680	nex =	1	
35	Sngl	42885	43564	+	0
	>2832689 /34867				
	len =	2058	nex =	6	
45	Init	65633	65767	+	0
	Intr	65857	66031	+	0
	Intr	66109	66211	+	0
	Intr	66295	66522	+	0
	Intr	66616	66834	+	0
50	Term	66931	67577	+	0
	>2832689 /26737				
	len =	1390	nex =	2	
55	Init	6827	6884	+	0
	Term	6976	8196	+	0
	>2833627 /38715				
60					

2025-09-10 10:50:00

1051

	len =	4289	nex =	8	
	Init	18099	18163	+	0
	Intr	18434	18574	+	0
5	Intr	18921	18969	+	0
	Intr	19215	19331	+	0
	Intr	19412	19465	+	0
	Intr	19862	19950	+	0
	Intr	20099	20204	+	0
10	Term	20282	20669	+	0
>2833627		/30652			
	len =	2745	nex =	8	
15	Init	18099	18163	+	0
	Intr	18434	18574	+	0
	Intr	18921	18969	+	0
	Intr	19215	19331	+	0
20	Intr	19412	19465	+	0
	Intr	19862	19950	+	0
	Intr	20099	20204	+	0
	Term	20282	20669	+	0
25	>2833627	/38934			
	len =	2757	nex =	8	
	Init	18099	18163	+	0
30	Intr	18434	18574	+	0
	Intr	18921	18969	+	0
	Intr	19215	19331	+	0
	Intr	19412	19465	+	0
	Intr	19862	19950	+	0
35	Intr	20099	20204	+	0
	Term	20282	20681	+	0
>2833627		/29043			
40	len =	2111	nex =	8	
	Init	67181	67444	+	0
	Intr	67542	67618	+	0
	Intr	67706	67828	+	0
45	Intr	68112	68216	+	0
	Intr	68331	68403	+	0
	Intr	68482	68749	+	0
	Intr	68837	68966	+	0
	Term	69054	69291	+	0
50	>2842474	/10704			
	len =	825	nex =	3	
55	Init	21629	21886	+	0
	Intr	21961	21999	+	0
	Term	22252	22453	+	0
>2842474		/5117			
60					

2025-09-23 10:00:00

1052

	len =	734	nex =	1	
	Sngl	23867	24227	+	0
5	>2842474	/1885			
	len =	2068	nex =	7	
	Init	35358	35949	+	0
10	Intr	36030	36183	+	0
	Intr	36287	36350	+	0
	Intr	36434	36537	+	0
	Intr	36679	36748	+	0
	Intr	36860	36925	+	0
15	Term	37010	37425	+	0
	>2842474	/41008			
	len =	1281	nex =	2	
20	Init	37554	37887	+	0
	Term	38343	38834	+	0
	>2842474	/906			
25	len =	3084	nex =	16	
	Term	39452	39140	-	0
	Intr	39647	39559	-	0
30	Intr	39923	39818	-	0
	Intr	40163	40038	-	0
	Intr	40346	40261	-	0
	Intr	40503	40450	-	0
	Intr	40671	40608	-	0
35	Intr	40868	40773	-	0
	Intr	41029	40955	-	0
	Intr	41151	41117	-	0
	Intr	41302	41233	-	0
	Intr	41481	41383	-	0
40	Intr	41624	41597	-	0
	Intr	41822	41711	-	0
	Intr	42001	41911	-	0
	Init	42223	42084	-	0
45	>2842474	/40688			
	len =	2028	nex =	5	
	Term	62730	62375	-	0
50	Intr	63029	62817	-	0
	Intr	63288	63118	-	0
	Intr	63523	63407	-	0
	Init	63946	63768	-	0
55	>2842474	/35774			
	len =	4104	nex =	3	
	Term	68866	68144	-	0
60	Intr	70164	69899	-	0

1053

	Init	72247	71953	-	0
	>2842474		/18360		
5	len =	1882	nex =	6	
	Term	7149	6711	-	0
	Intr	7371	7305	-	0
	Intr	7647	7537	-	0
10	Intr	7915	7763	-	0
	Intr	8238	8171	-	0
	Init	8592	8366	-	0
	>2853071		/19264		
15	len =	2650	nex =	3	
	Init	28721	29386	+	0
	Intr	29475	30150	+	0
20	Term	30940	31017	+	0
	>2853071		/42503		
	len =	1695	nex =	1	
25	Sngl	4329	2635	-	0
	>2853071		/13295		
30	len =	1851	nex =	8	
	Init	49706	49914	+	0
	Intr	50003	50111	+	0
	Intr	50216	50338	+	0
35	Intr	50428	50525	+	0
	Intr	50616	50742	+	0
	Intr	50822	50907	+	0
	Intr	50994	51119	+	0
	Term	51220	51556	+	0
40	>2853071		/24087		
	len =	799	nex =	1	
45	Sngl	74493	75291	+	0
	>2853071		/143232		
	len =	689	nex =	2	
50	Term	76634	76317	-	0
	Init	77005	76728	-	0
	>2864607		/34819		
55	len =	1657	nex =	3	
	Init	11411	11849	+	0
	Intr	12198	12373	+	0
60	Term	12788	13067	+	0

2025-01-07 14:59:59

	>2864607	/31896		
5	len =	490	nex =	1
	Sngl	35491	35976	+ 0
	>2864607	/20798		
10	len =	1016	nex =	2
	Init	50921	51566	+ 0
	Term	51635	51936	+ 0
15	>2864607	/30539		
	len =	995	nex =	2
20	Init	50974	51566	+ 0
	Term	51635	51968	+ 0
	>2864607	/30310		
25	len =	779	nex =	3
	Term	56737	56505	- 0
	Intr	56871	56822	- 0
	Init	57157	56983	- 0
30	>2864607	/29964		
	len =	877	nex =	4
	Term	56737	56470	- 0
35	Intr	56871	56822	- 0
	Intr	57157	56983	- 0
	Init	57346	57251	- 0
	>2864607	/115583		
40	len =	1150	nex =	1
	Sngl	91469	90322	- 0
45	>2880038	/39751		
	len =	2230	nex =	7
	Term	33273	33017	- 0
50	Intr	33540	33439	- 0
	Intr	33821	33808	- 0
	Intr	34285	34177	- 0
	Intr	34463	34390	- 0
	Intr	34593	34550	- 0
55	Init	34847	34762	- 0
	>2880038	/20095		
60	len =	2306	nex =	8

2025-09-23 10:54:00

1055

5	Term	33273	32976	-	0
	Intr	33540	33439	-	0
	Intr	33821	33808	-	0
	Intr	34285	34177	-	0
	Intr	34463	34390	-	0
	Intr	34593	34550	-	0
	Intr	34847	34762	-	0
	Init	35271	35063	-	0
10	>2880038	/31304			
	len =	310	nex =	1	
15	Sngl	75	382	+	0
	>2894557	/98545			
	len =	953	nex =	2	
20	Term	11319	10903	-	0
	Init	11855	11346	-	0
	>2894557	/14645			
25	len =	1007	nex =	3	
	Term	11187	10877	-	0
	Intr	11502	11346	-	0
30	Init	11883	11602	-	0
	>2894557	/14462			
	len =	1180	nex =	5	
35	Term	13853	13556	-	0
	Intr	14042	14002	-	0
	Intr	14172	14109	-	0
	Intr	14432	14276	-	0
40	Init	14735	14527	-	0
	>2894557	/24573			
	len =	1059	nex =	3	
45	Term	17997	17644	-	0
	Intr	18258	18102	-	0
	Init	18702	18484	-	0
	>2894557	/16643			
50	len =	1411	nex =	3	
	Term	26860	26567	-	0
	Intr	27098	26942	-	0
55	Init	27977	27679	-	0
	>2894557	/1004			
	len =	1369	nex =	3	
60					

1056

Init	28664	28756	+	0
Intr	29030	29218	+	0
Term	29365	29739	+	0

5 >2894557 /4764

len =	610	nex =	1	
Sngl	30727	30125	-	0

10 >2894557 /23223

len =	670	nex =	2	
Term	30491	30129	-	0
Init	30795	30617	-	0

>2894557 /28411

20 len = 1931 nex = 6

Term	1991	1564	-	0
Intr	2137	2084	-	0
Intr	2311	2228	-	0
Intr	2623	2409	-	0
Init	3127	3028	-	0
Init	3494	3369	-	0

>2894557 /17340

30 len = 5636 nex = 8

Term	52206	51907	-	0
Intr	52347	52286	-	0
Intr	52539	52453	-	0
Intr	52749	52631	-	0
Intr	53083	52845	-	0
Intr	53233	53177	-	0
Intr	54012	53381	-	0
Init	57542	57102	-	0

>2894557 /32972

45 len = 328 nex = 1

Sngl	73075	72748	-	0
------	-------	-------	---	---

>2894557 /33172

50 len = 2020 nex = 4

Term	73486	72801	-	0
Intr	74137	73972	-	0
Intr	74427	74239	-	0
Init	74820	74549	-	0

>2894557 /20758

60 len = 1959 nex = 4

1057

	Term	73486	72863	-	0
	Intr	74137	73972	-	0
	Intr	74427	74239	-	0
	Init	74821	74549	-	0
5	>2894591 /37898				
	len =	1361	nex =	7	
10	Init	4	48	+	0
	Intr	125	263	+	0
	Intr	346	421	+	0
	Intr	502	624	+	0
	Intr	715	868	+	0
15	Intr	975	1008	+	0
	Term	1090	1361	+	0
	>2894591 /4613				
20	len =	818	nex =	1	
	Sngl	11818	12635	+	0
	>2894591 /12013				
25	len =	2086	nex =	3	
	Init	29017	29255	+	0
	Intr	29840	30037	+	0
30	Term	30126	30344	+	0
	>2894591 /38412				
35	len =	2218	nex =	10	
	Term	1518	1201	-	0
	Intr	1688	1600	-	0
	Intr	1839	1768	-	0
	Intr	2024	1922	-	0
40	Intr	2182	2114	-	0
	Intr	2336	2274	-	0
	Intr	2511	2416	-	0
	Intr	2749	2598	-	0
	Intr	3311	2863	-	0
45	Init	3418	3345	-	0
	>2894591 /33385				
	len =	1224	nex =	3	
50	Init	48888	49463	+	0
	Intr	49539	49711	+	0
	Term	49804	50111	+	0
55	>2894591 /13507				
	len =	1610	nex =	2	
	Init	53993	54524	+	0
60	Term	55134	55602	+	0

2025-09-26 10:00

	>2914688	/27602		
5	len =	827	nex =	2
	Init	32845	33192	+
	Term	33282	33671	+
				0
10	>2914688	/27570		
	len =	500	nex =	1
	Sngl	3483	3969	+
				0
15	>2914688	/30472		
	len =	650	nex =	2
20	Term	43833	43650	-
	Init	44299	43939	-
				0
	>2914688	/39917		
25	len =	1453	nex =	2
	Term	63880	62989	-
	Init	64441	63995	-
				0
30	>2914688	/7865		
	len =	398	nex =	1
	Sngl	74566	74860	+
				0
35	>2914688	/154660		
	len =	406	nex =	1
40	Sngl	75023	75428	+
				0
	>2914688	/3956		
	len =	380	nex =	1
45	Sngl	75051	75430	+
				0
	>2914688	/146794		
50	len =	202	nex =	1
	Sngl	75123	75324	+
				0
	>2914688	/23742		
55	len =	987	nex =	1
	Sngl	87641	87085	-
				0
60	>2914688	/28031		

1060

	Term	65652	65912	+	0
	>2924651	/116144			
5	len =	550	nex =	1	
	Sngl	11422	11968	+	0
	>2924651	/4228			
10	len =	1336	nex =	1	
	Sngl	16008	16108	+	0
15	>2924651	/41343			
	len =	1397	nex =	2	
	Init	16012	16108	+	0
20	Term	16483	17408	+	0
	>2924651	/37811			
	len =	3460	nex =	14	
25	Term	17854	17679	-	0
	Intr	17949	17883	-	0
	Intr	18132	18040	-	0
	Intr	18319	18226	-	0
30	Intr	18472	18400	-	0
	Intr	19189	19124	-	0
	Intr	19363	19286	-	0
	Intr	19517	19431	-	0
35	Intr	20061	19866	-	0
	Intr	20202	20146	-	0
	Intr	20333	20276	-	0
	Intr	20770	20446	-	0
	Intr	20948	20865	-	0
40	Init	21138	21036	-	0
	>2924651	/11706			
	len =	797	nex =	4	
45	Init	27009	27125	+	0
	Intr	27212	27285	+	0
	Intr	27349	27473	+	0
	Term	27545	27805	+	0
50	>2924651	/101608			
	len =	798	nex =	3	
	Init	27010	27125	+	0
55	Intr	27212	27473	+	0
	Term	27545	27807	+	0
	>2924651	/11215			
60	len =	1990	nex =	4	

2025-01-06 10:00:00

1061

	Term	40287	39899	-	0
	Intr	40599	40363	-	0
	Intr	41110	40743	-	0
5	Init	41882	41190	-	0
	>2924651		/124621		
	len =	1394	nex =	5	
10	Init	42248	42476	+	0
	Intr	42705	42841	+	0
	Intr	42918	42975	+	0
	Intr	43069	43146	+	0
15	Term	43242	43641	+	0
	>2924651		/115662		
	len =	1586	nex =	3	
20	Term	54089	53651	-	0
	Intr	54599	54185	-	0
	Init	55236	54985	-	0
25	>2924651		/24140		
	len =	989	nex =	5	
	Init	64216	64414	+	0
30	Intr	64498	64577	+	0
	Intr	64656	64726	+	0
	Intr	64819	64872	+	0
	Term	64968	65204	+	0
35	>2924651		/41440		
	len =	520	nex =	2	
	Term	67750	67536	-	0
40	Init	68055	67838	-	0
	>2924651		/27491		
	len =	1390	nex =	7	
45	Term	68218	68177	-	0
	Intr	68390	68313	-	0
	Intr	68583	68492	-	0
	Intr	68745	68667	-	0
50	Intr	68963	68830	-	0
	Intr	69280	69052	-	0
	Init	69566	69388	-	0
	>2924651		/18901		
55	len =	2963	nex =	12	
	Term	73319	73016	-	0
	Intr	73588	73415	-	0
60	Intr	73756	73694	-	0

2025 RELEASE UNDER E.O. 14176

1062

	Intr	73960	73853	-	0
	Intr	74160	74045	-	0
	Intr	74372	74246	-	0
	Intr	74635	74466	-	0
5	Intr	74806	74728	-	0
	Intr	75007	74896	-	0
	Intr	75150	75062	-	0
	Intr	75362	75232	-	0
	Init	75978	75624	-	0
10	>2924651 /111348				
	len =	681	nex =	2	
15	Init	79079	79259	+	0
	Term	79349	79759	+	0
	>2924652 /4911				
20	len =	1030	nex =	3	
	Term	19435	19201	-	0
	Intr	19824	19635	-	0
	Init	20222	20090	-	0
25	>2924652 /36891				
	len =	1540	nex =	3	
30	Term	34140	33486	-	0
	Intr	34498	34229	-	0
	Init	35025	34629	-	0
35	>2924652 /6935				
	len =	1820	nex =	3	
	Term	34140	33312	-	0
	Intr	34498	34229	-	0
40	Init	35131	34629	-	0
	>2924653 /24054				
	len =	985	nex =	1	
45	Sngl	18725	19123	+	0
	>2924653 /40303				
50	len =	2207	nex =	7	
	Init	48858	49136	+	0
	Intr	49281	49579	+	0
	Intr	49695	49920	+	0
55	Intr	50013	50258	+	0
	Intr	50352	50462	+	0
	Intr	50538	50669	+	0
	Term	50749	51064	+	0
60	>2924653 /37280				

1063

	len =	2358	nex =	10	
	Init	5790	5865	+	0
5	Intr	6335	6391	+	0
	Intr	6473	6632	+	0
	Intr	6769	6851	+	0
	Intr	6950	7009	+	0
	Intr	7106	7214	+	0
10	Intr	7290	7342	+	0
	Intr	7437	7517	+	0
	Intr	7595	7707	+	0
	Term	7811	8147	+	0
15	>2924653 /10022				
	len =	2002	nex =	4	
	Init	64053	64145	+	0
20	Intr	64317	64444	+	0
	Intr	64584	64688	+	0
	Term	65141	65503	+	0
	>2924653 /12547				
25	len =	2003	nex =	4	
	Init	64053	64145	+	0
	Intr	64317	64444	+	0
30	Intr	64584	64688	+	0
	Term	65141	65504	+	0
	>2924653 /6841				
35	len =	1345	nex =	3	
	Init	70235	70473	+	0
	Intr	70569	70632	+	0
	Term	71064	71579	+	0
40	>2924653 /22448				
	len =	1249	nex =	3	
45	Init	70377	70473	+	0
	Intr	70569	70632	+	0
	Term	71064	71625	+	0
	>2924653 /92808				
50	len =	850	nex =	3	
	Term	74646	74511	-	0
	Intr	74845	74760	-	0
55	Init	75351	74975	-	0
	>2924654 /30054				
60	len =	2574	nex =	7	

1064

	Init	14617	14705	+	0
	Intr	15035	15246	+	0
	Intr	15355	15634	+	0
	Intr	15717	15788	+	0
5	Intr	15879	15995	+	0
	Intr	16116	16181	+	0
	Term	16657	16957	+	0
	>2924654		/37363		
10	len =	2618	nex =	6	
	Init	14617	15246	+	0
	Intr	15355	15634	+	0
15	Intr	15717	15788	+	0
	Intr	15879	15995	+	0
	Intr	16116	16181	+	0
	Term	16657	17013	+	0
20	>2924654		/41320		
	len =	2533	nex =	7	
	Init	14422	14705	+	0
25	Intr	15035	15246	+	0
	Intr	15355	15634	+	0
	Intr	15717	15788	+	0
	Intr	15879	15995	+	0
	Intr	16116	16181	+	0
30	Term	16657	16954	+	0
	>2924654		/105970		
	len =	997	nex =	1	
35					
	Sngl	29862	28866	-	0
	>2924654		/37399		
40	len =	2110	nex =	5	
	Init	41781	42184	+	0
	Intr	42289	42366	+	0
	Intr	42453	42522	+	0
45	Intr	42650	42927	+	0
	Term	43021	43108	+	0
	>2924655		/33830		
50	len =	2316	nex =	10	
	Term	18451	18193	-	0
	Intr	18825	18612	-	0
	Intr	19016	18919	-	0
55	Intr	19270	19142	-	0
	Intr	19331	19289	-	0
	Intr	19690	19580	-	0
	Intr	19925	19779	-	0
	Intr	20137	20024	-	0
60	Intr	20329	20200	-	0

1065

	Init	20508	20379	-	0
	>2924655	/37918			
5	len =	799	nex =	3	
	Init	34127	34257	+	0
	Intr	34452	34520	+	0
	Term	34618	34925	+	0
10	>2924728	/112173			
	len =	321	nex =	1	
15	Sngl	16725	16405	-	0
	>2924728	/20919			
	len =	3224	nex =	12	
20	Init	39629	39916	+	0
	Intr	40038	40180	+	0
	Intr	40259	40351	+	0
	Intr	40438	40554	+	0
25	Intr	40669	40764	+	0
	Intr	40904	41017	+	0
	Intr	41217	41270	+	0
	Intr	41480	41527	+	0
	Intr	41825	41975	+	0
30	Intr	42080	42245	+	0
	Intr	42327	42396	+	0
	Term	42476	42852	+	0
	>2924728	/15599			
35	len =	290	nex =	1	
	Sngl	46957	46668	-	0
40	>2924728	/32450			
	len =	2230	nex =	2	
	Term	47403	46673	-	0
45	Init	48900	48211	-	0
	>2924728	/11304			
	len =	1048	nex =	1	
50	Sngl	4497	3866	-	0
	>2924728	/11830			
55	len =	740	nex =	3	
	Term	56509	56324	-	0
	Intr	56849	56595	-	0
	Init	57063	56929	-	0
60					

	>2924729	/32977		
	len =	597	nex =	1
5	Sngl	32483	31887	- 0
	>2924729	/8150		
	len =	1108	nex =	3
10	Term	37394	37188	- 0
	Intr	37888	37601	- 0
	Init	38295	37976	- 0
15	>2924729	/108595		
	len =	698	nex =	1
20	Sngl	41180	40483	- 0
	>2924729	/27221		
	len =	1539	nex =	3
25	Term	60702	60586	- 0
	Intr	61397	61275	- 0
	Init	61782	61669	- 0
	>2924729	/12817		
30	len =	1994	nex =	3
	Term	60702	60586	- 0
	Intr	61397	61275	- 0
35	Init	62241	61669	- 0
	>2924730	/14419		
	len =	559	nex =	1
40	Sngl	3605	3059	- 0
	>2924730	/9214		
45	len =	328	nex =	1
	Sngl	8331	8658	+ 0
	>2924731	/34161		
50	len =	2210	nex =	8
	Init	3351	3568	+ 0
	Intr	3663	3747	+ 0
55	Intr	3839	3906	+ 0
	Intr	4037	4087	+ 0
	Intr	4172	4245	+ 0
	Intr	4652	4710	+ 0
	Intr	4794	4870	+ 0
60	Term	5021	5301	+ 0

2025 RELEASE UNDER E.O. 14176

1067

>2924732 /23726

5	len =	2127	nex =	6	
	Term	15136	14749	-	0
	Intr	15419	15211	-	0
	Intr	15598	15512	-	0
	Intr	15788	15678	-	0
10	Intr	16145	16078	-	0
	Init	16299	16228	-	0

>2924732 /118484

15	len =	1335	nex =	2	
	Term	17876	17178	-	0
	Init	18512	18040	-	0

20 >2924732 /110980

	len =	1726	nex =	6	
	Term	2588	2145	-	0
25	Intr	2744	2665	-	0
	Intr	2992	2845	-	0
	Intr	3155	3095	-	0
	Intr	3489	3265	-	0
	Init	3870	3763	-	0

>2924732 /10411

	len =	770	nex =	1	
35	Sngl	43741	42972	-	0

>2924732 /19177

40	len =	829	nex =	1	
	Sngl	45210	44382	-	0

>2924732 /37398

45	len =	2350	nex =	9	
	Term	72077	71790	-	0
	Intr	72317	72183	-	0
	Intr	72514	72395	-	0
50	Intr	72722	72612	-	0
	Intr	73197	72982	-	0
	Intr	73331	73278	-	0
	Intr	73501	73430	-	0
	Intr	73771	73595	-	0
55	Init	74137	73854	-	0

>2924733 /12093

60	len =	573	nex =	2	
----	-------	-----	-------	---	--

2025 RELEASE UNDER E.O. 14176

1068

Term	12077	11784	-	0
Init	12356	12172	-	0

>2924733 /30414

5

```
len =      397      nex =      2
```

Init	2448	2519	+	0
Term	2615	2844	+	0

10

>2924733 /147492

```
len =      621    nex =      1
```

15	Snq1	38573	37953	-	0
----	------	-------	-------	---	---

>2924733 / 8989

```
len = 2384    nex = 8
```

20

Init	465	737	+	0
------	-----	-----	---	---

Intr	820	941	+	0
------	-----	-----	---	---

Intr	1362	1477	+	0
------	------	------	---	---

Intr	1616	1689	+	0
------	------	------	---	---

25

Intr	1800	2064	+	0
------	------	------	---	---

Intr	2148	2279	+	0
------	------	------	---	---

Intr	2384	2519	+	0
------	------	------	---	---

Term	2615	2848	+	0
------	------	------	---	---

30 >2924733 /20896

```
len = 2392    nex = 8
```

	Init	465	737	+	0
35	Inter	820	941	+	0

Incl	820	941	+	0
Intr	1362	1477	+	0

Incl	1502	1477	+	0
Intr	1616	1689	+	0

Intr	1810	1889	+	0
Intr	1800	2064	+	0

Incl	1800	2004	-	0
Intr	2148	2279	+	0

Inter	2140	2279	+	0
Intr	2384	2519	+	0

40

Term	2615	2856	+	0
------	------	------	---	---

>2924733 /25282

```
45      len =      2410      nex =      8
```

Init	465	737	+	0
------	-----	-----	---	---

Intr	820	941	+	0
------	-----	-----	---	---

Intr	1362	1477	+	0
------	------	------	---	---

50

Intr	1616	1689	+	0
------	------	------	---	---

Intr	1800	2064	+	0
------	------	------	---	---

Intr	2148	2279	+	0
------	------	------	---	---

Intr	2384	2519	+	0
------	------	------	---	---

55

>2924733 /15239

```
len = 976    nex = 1
```

60	Snq1	47202	46227	-	(
----	------	-------	-------	---	---

1070

>2924768 /2443

len = 2009 nex = 5

5	Init	54917	55312	+	0
	Intr	55789	55930	+	0
	Intr	56024	56144	+	0
	Intr	56337	56399	+	0
	Term	56526	56925	+	0

10

>2924768 /40107

len = 1793 nex = 2

15	Init	55097	55930	+	0
	Term	56024	56889	+	0

>2947056 /35441

20 len = 1884 nex = 7

	Term	24625	24082	-	0
	Intr	24833	24712	-	0
	Intr	25012	24909	-	0
25	Intr	25146	25116	-	0
	Intr	25374	25244	-	0
	Intr	25569	25472	-	0
	Init	25953	25760	-	0

30 >2947056 /18146

len = 884 nex = 5

	Term	52485	52355	-	0
35	Intr	52728	52640	-	0
	Intr	52902	52813	-	0
	Intr	53019	52994	-	0
	Init	53238	53111	-	0

40 >2961335 /19666

len = 970 nex = 3

	Init	100334	100454	+	0
45	Intr	100557	100713	+	0
	Term	100813	101289	+	0

>2961335 /42504

50 len = 1641 nex = 4

	Term	101428	101392	-	0
	Intr	102385	102247	-	0
	Intr	102589	102521	-	0
55	Init	103032	102688	-	0

>2961335 /37217

60 len = 1635 nex = 4

1071

	Term	101428	101398	-	0
	Intr	102385	102247	-	0
	Intr	102589	102521	-	0
	Init	103032	102688	-	0
5	>2961335 /39963				
	len =	889	nex =	1	
10	Sngl	31532	30644	-	0
	>2961335 /36866				
	len =	2013	nex =	4	
15	Term	31944	30671	-	0
	Intr	32165	32030	-	0
	Intr	32386	32252	-	0
	Init	32683	32529	-	0
20	>2961335 /21244				
	len =	1212	nex =	1	
25	Sngl	33841	33295	-	0
	>2961335 /322				
	len =	2495	nex =	9	
30	Term	85172	85058	-	0
	Intr	85456	85295	-	0
	Intr	85703	85608	-	0
	Intr	85906	85842	-	0
35	Intr	86542	86467	-	0
	Intr	86714	86632	-	0
	Intr	87177	86849	-	0
	Intr	87304	87258	-	0
	Init	87546	87386	-	0
40	>2961370 /14792				
	len =	1218	nex =	2	
45	Init	2148	2639	+	0
	Term	2907	3365	+	0
	>2961370 /41090				
50	len =	271	nex =	1	
	Sngl	3095	3365	+	0
	>2961370 /36110				
55	len =	2782	nex =	12	
	Init	31733	31968	+	0
	Intr	32075	32200	+	0
60	Intr	32291	32450	+	0

2025 RELEASE UNDER E.O. 14176

1072

	Intr	32544	32729	+	0
	Intr	32831	32947	+	0
	Intr	33047	33100	+	0
5	Intr	33180	33251	+	0
	Intr	33395	33579	+	0
	Intr	33685	33779	+	0
	Intr	33862	33936	+	0
	Intr	34061	34144	+	0
	Term	34259	34514	+	0
10	>2961370 /9402				
	len =	329	nex =	1	
15	Sngl	52923	53251	+	0
	>2961370 /12881				
	len =	375	nex =	1	
20	Sngl	88516	88142	-	0
	>2979540 /12830				
25	len =	1150	nex =	2	
	Term	31157	30569	-	0
	Init	31718	31304	-	0
30	>2979540 /11908				
	len =	238	nex =	1	
	Sngl	43116	43353	+	0
35	>2979540 /9990				
	len =	1645	nex =	4	
40	Term	53451	53377	-	0
	Intr	53944	53849	-	0
	Intr	54297	54187	-	0
	Init	54719	54557	-	0
45	>2979540 /5467				
	len =	3267	nex =	7	
	Term	61277	60884	-	0
50	Intr	61544	61503	-	0
	Intr	61708	61667	-	0
	Intr	61911	61812	-	0
	Intr	62080	62019	-	0
	Intr	62353	62272	-	0
55	Init	63560	63356	-	0
	>2979540 /100354				
60	len =	1123	nex =	3	

1073

Term	68542	68068	-	0
Intr	68676	68629	-	0
Init	69047	68997	-	0

5 >2979540 /18662

len = 1011 nex = 3

Init	84864	84965	+	0
Intr	85100	85128	+	0
Term	85223	85515	+	0

>2979540 /9848

15 len = 478 nex = 3

Init	84863	84965	+	0
Intr	85100	85128	+	0
Term	85223	85340	+	0

20 >2979540 /123760

len = 631 nex = 3

Term	9120	9002	-	0
Intr	9356	9289	-	0
Init	9632	9459	-	0

>2979540 /35578

30 len = 976 nex = 1

Sngl 9822 10797 + 0

35 >2980757 /96478

len = 628 nex = 2

Term	101920	101603	-	0
Init	102230	102001	-	0

>2980757 /29888

45 len = 1706 nex = 2

Init	105420	106491	+	0
Term	106578	106835	+	0

>2980757 /20759

50 len = 1674 nex = 1

Sngl 111308 109635 - 0

55 >2980757 /32640

len = 1975 nex = 5

Term	113953	113573	-	0
Intr	114107	114042	-	0

60

1074

Intr	114352	114249	-	0
Intr	114505	114432	-	0
Init	115547	115405	-	0

5 >2980757 /14193

len = 2050 nex = 4

Term	113953	113573	-	0
Intr	114107	114042	-	0
Intr	114352	114249	-	0
Init	114505	114432	-	0

>2980757 /39459

15 len = 2077 nex = 4

Term	113953	113570	-	0
Intr	114107	114042	-	0
Intr	114352	114249	-	0
Init	114505	114432	-	0

>2980757 /20888

25 len = 3039 nex = 9

Term	116278	115618	-	0
Intr	116459	116365	-	0
Intr	116906	116773	-	0
Intr	117244	116981	-	0
Intr	117420	117335	-	0
Intr	117593	117501	-	0
Intr	117776	117681	-	0
Intr	117973	117866	-	0
Init	118656	118366	-	0

>2980757 /17828

40 len = 1390 nex = 3

Term	19947	19431	-	0
Intr	20153	20130	-	0
Init	20797	20451	-	0

45 >2980757 /37878

len = 1891 nex = 7

Term	22249	22095	-	0
Intr	22401	22353	-	0
Intr	22603	22491	-	0
Intr	22808	22697	-	0
Intr	23041	22956	-	0
Intr	23548	23465	-	0
Init	23985	23709	-	0

>2980757 /40278

60 len = 2700 nex = 10

1075

	Init	62319	62472	+	0
	Intr	62599	62662	+	0
	Intr	62935	63182	+	0
	Intr	63261	63404	+	0
5	Intr	63478	63629	+	0
	Intr	63712	63834	+	0
	Intr	63933	64110	+	0
	Intr	64204	64322	+	0
	Intr	64409	64523	+	0
10	Term	64665	65018	+	0
>2980757 /7876					
	len =	1620	nex =	5	
15	Init	70442	70911	+	0
	Intr	70996	71076	+	0
	Intr	71232	71333	+	0
	Intr	71626	71700	+	0
20	Term	71818	72061	+	0
>2980757 /31287					
	len =	2424	nex =	7	
25	Init	8024	8864	+	0
	Intr	9004	9089	+	0
	Intr	9190	9363	+	0
	Intr	9500	9614	+	0
30	Intr	9732	9829	+	0
	Intr	9974	10109	+	0
	Term	10228	10447	+	0
>2980787 /27462					
35	len =	2028	nex =	4	
	Term	46006	45949	-	0
	Intr	46321	46080	-	0
40	Intr	46522	46410	-	0
	Init	47415	47305	-	0
>3004543 /39073					
45	len =	1651	nex =	5	
	Init	15971	16170	+	0
	Intr	16291	16381	+	0
	Intr	16472	17062	+	0
50	Intr	17134	17233	+	0
	Term	17529	17621	+	0
>3004543 /15729					
55	len =	1699	nex =	5	
	Init	15980	16150	+	0
	Intr	16291	16381	+	0
	Intr	16472	17062	+	0
60	Intr	17134	17233	+	0

2025-09-23 10:00:00

1076

	Term	17529	17678	+	0
	>3004543	/1314			
5	len =	1330	nex =	5	
	Init	21535	21572	+	0
	Intr	21645	21756	+	0
	Intr	21837	21934	+	0
10	Intr	22311	22407	+	0
	Term	22482	22571	+	0
	>3004543	/3350			
15	len =	1450	nex =	3	
	Term	44799	44453	-	0
	Intr	45325	45228	-	0
	Init	45894	45618	-	0
20	>3004543	/36487			
	len =	770	nex =	2	
25	Term	45325	45258	-	0
	Init	46027	45618	-	0
	>3004543	/16223			
30	len =	1140	nex =	4	
	Init	78104	78275	+	0
	Intr	78360	78542	+	0
	Intr	78629	78810	+	0
35	Term	78902	79243	+	0
	>3021263	/1667			
40	len =	868	nex =	1	
	Sngl	23858	22991	-	0
	>3033373	/21342			
45	len =	747	nex =	1	
	Sngl	15077	15823	+	0
	>3033373	/29605			
50	len =	595	nex =	1	
	Sngl	39454	40045	+	0
55	>3033373	/6606			
	len =	1090	nex =	1	
	Sngl	79964	78877	-	0
60					

2025-08-16 09:59:59

1077

>3033373 /24400

```
len = 2003    nex = 6
```

5	Init	82999	83318	+	0
	Intr	83637	83702	+	0
	Intr	83796	83846	+	0
	Intr	83938	84017	+	0
	Intr	84308	84352	+	0
10	Term	84467	85001	+	0

>3033373 /5509

```
len = 3955    nex = 15
```

15	len	5599	len	15
	Init	85397	85674	+
	Intr	85852	85906	+
	Intr	85992	86143	+
	Intr	86819	86858	+
20	Intr	87036	87128	+
	Intr	87213	87328	+
	Intr	87417	87491	+
	Intr	87611	87650	+
	Intr	87739	87821	+
25	Intr	87911	87988	+
	Intr	88122	88265	+
	Intr	88458	88525	+
	Intr	88608	88722	+
	Intr	88830	88986	+
30	Term	89057	89351	+

>3036791 /34369

```
len = 1935    nex = 7
```

	Term	30641	30380	-	0
	Intr	30794	30729	-	0
	Intr	30936	30872	-	0
	Intr	31426	31363	-	0
40	Intr	31610	31501	-	0
	Intr	31784	31697	-	0
	Init	32314	31905	-	0

>3036791 /11954

```
len = 1033    nex = 1
```

Snq1	56760	55728	-	0
------	-------	-------	---	---

50 >3036791 /103458

```
len = 1150    nex = 1
```

Snq1	58836	59073	+	0
------	-------	-------	---	---

55 >3036791 /4329

```
len = 1227    nex = 2
```

```
60      Init    90374    91016      +      0
```

1078

	Term	91266	91600	+	0
	>3036791	/14578			
5	len =	880	nex =	1	
	Sngl	96436	97315	+	0
	>3046847	/8184			
10	len =	3010	nex =	10	
	Term	8682	8425	-	0
	Intr	8887	8808	-	0
15	Intr	9087	8960	-	0
	Intr	9281	9216	-	0
	Intr	9673	9622	-	0
	Intr	9870	9785	-	0
	Intr	10314	10151	-	0
20	Intr	10511	10395	-	0
	Intr	10668	10598	-	0
	Init	11428	11126	-	0
	>3046847	/37197			
25	len =	1990	nex =	5	
	Term	11964	11553	-	0
	Intr	12305	12048	-	0
30	Intr	12692	12466	-	0
	Intr	12917	12776	-	0
	Init	13535	13263	-	0
	>3046847	/19893			
35	len =	858	nex =	1	
	Sngl	39899	39042	-	0
40	>3046847	/92102			
	len =	992	nex =	1	
	Sngl	42005	41014	-	0
45	>3046847	/2037			
	len =	585	nex =	2	
50	Term	43375	42957	-	0
	Init	43541	43466	-	0
	>3046847	/95135			
55	len =	1247	nex =	2	
	Init	49610	50038	+	0
	Term	50471	50856	+	0
60	>3046847	/9324			

2025 RELEASE UNDER E.O. 14176

1079

5 len = 479 nex = 1
 Sngl 54499 54021 - 0
 >3046849 /14126
 len = 1577 nex = 2
 10 Term 28105 27242 - 0
 Init 28795 28556 - 0
 >3046849 /156665
 15 len = 839 nex = 2
 Term 28105 27980 - 0
 Init 28795 28556 - 0
 20 >3046849 /3963
 len = 1488 nex = 2
 Term 39835 39089 - 0
 25 Init 40576 40199 - 0
 >3046849 /13314
 len = 1665 nex = 5
 30 Init 44226 44488 + 0
 Intr 44558 44738 + 0
 Intr 44824 44990 + 0
 Intr 45085 45167 + 0
 35 Term 45367 45890 + 0
 >3046849 /40824
 len = 1361 nex = 5
 40 Init 44387 44488 + 0
 Intr 44558 44738 + 0
 Intr 44824 44990 + 0
 Intr 45085 45167 + 0
 45 Term 45367 45747 + 0
 >3046850 /32178
 len = 1227 nex = 5
 50 Term 49 1 - 0
 Intr 218 144 - 0
 Intr 570 392 - 0
 Intr 748 681 - 0
 55 Init 1227 972 - 0
 >3046850 /23581
 60 len = 1553 nex = 4

1080

	Term	26495	26176	-	0
	Intr	27046	26959	-	0
	Intr	27255	27139	-	0
	Init	27728	27452	-	0
5	>3046850 /4275				
	len =	1570	nex =	4	
10	Term	26495	26189	-	0
	Intr	27046	26959	-	0
	Intr	27255	27139	-	0
	Init	27752	27452	-	0
15	>3046850 /18867				
	len =	2066	nex =	6	
20	Term	35882	35690	-	0
	Intr	36071	35971	-	0
	Intr	36216	36177	-	0
	Intr	36347	36300	-	0
	Intr	36722	36567	-	0
	Init	37087	36940	-	0
25	>3046850 /1680				
	len =	580	nex =	2	
30	Term	44554	44138	-	0
	Init	44695	44629	-	0
	>3046850 /27167				
35	len =	1215	nex =	3	
	Term	44554	44109	-	0
	Intr	44695	44629	-	0
	Init	45323	45131	-	0
40	>3046850 /2557				
	len =	790	nex =	3	
45	Term	58982	58595	-	0
	Intr	59222	59091	-	0
	Init	59377	59312	-	0
	>3046850 /21485				
50	len =	1210	nex =	4	
	Init	68376	68477	+	0
	Intr	68594	68717	+	0
55	Intr	69144	69216	+	0
	Term	69364	69581	+	0
	>3046850 /7347				
60	len =	1314	nex =	7	

2025 RELEASE UNDER E.O. 14176

1081

	Term	69704	69481	-	0
	Intr	69872	69789	-	0
	Intr	70063	69955	-	0
5	Intr	70270	70147	-	0
	Intr	70434	70338	-	0
	Intr	70647	70515	-	0
	Init	70794	70726	-	0
10	>3046850		/13780		
	len =	3435	nex =	11	
	Term	69704	69439	-	0
15	Intr	69872	69789	-	0
	Intr	70063	69955	-	0
	Intr	70270	70147	-	0
	Intr	70434	70338	-	0
	Intr	70647	70515	-	0
20	Intr	70803	70726	-	0
	Intr	70979	70872	-	0
	Intr	71133	71060	-	0
	Intr	71700	71616	-	0
	Init	71842	71778	-	0
25	>3046851		/39764		
	len =	2129	nex =	2	
30	Init	16125	17251	+	0
	Term	17913	18253	+	0
	>3046851		/19097		
35	len =	2129	nex =	2	
	Init	16148	17251	+	0
	Term	17913	18276	+	0
40	>3046851		/39024		
	len =	1270	nex =	2	
	Init	17002	17251	+	0
45	Term	17913	18259	+	0
	>3046851		/108127		
	len =	1035	nex =	2	
50	Init	17225	17251	+	0
	Term	17913	18259	+	0
	>3046851		/42919		
55	len =	2530	nex =	7	
	Init	23586	23813	+	0
	Intr	23901	24521	+	0
	Intr	24672	24859	+	0
60	Intr	25008	25246	+	0

2025 RELEASE UNDER E.O. 14176

1082

Intr	25330	25529	+	0
Intr	25634	25759	+	0
Term	25845	26114	+	0

5 >3046851 /34859

```
len =      1887      nex =          6
```

10	Init	24250	24521	+	0
	Intr	24672	24859	+	0
	Intr	25008	25246	+	0
	Intr	25330	25529	+	0
	Intr	25634	25759	+	0
	Term	25845	26136	+	0

15 >3046851 /103919

```
len =      918      nex =      1
```

20	Snq1	47374	48291	+	0
----	------	-------	-------	---	---

>3046852 /16392

```
len = 5153 nex = 13
```

Term	5264	5042	-	0
------	------	------	---	---

Intr	5566	5461	-	0
------	------	------	---	---

Intr	5753	5662	-	0
------	------	------	---	---

Intr	6129	6021	-	0
------	------	------	---	---

30	Intr	6633	6544	-	0
----	------	------	------	---	---

Intr	7388	7136	-	0
------	------	------	---	---

Intr	7603	7476	-	0
------	------	------	---	---

Intr	7755	7696	-	0
------	------	------	---	---

Intr	8145	8089	-	0
------	------	------	---	---

Intr	8399	8355	-	0
------	------	------	---	---

Intr	8654	8573	-	0
------	------	------	---	---

Intr	8799	8759	-	0
------	------	------	---	---

40 >3046852 /30518

```
len = 1038    nex = 3
```

Init	42931	43061	+	0
------	-------	-------	---	---

Intr	43292	43357	+	0
------	-------	-------	---	---

Term	43687	43968	+	0
------	-------	-------	---	---

>3046853 /36632

```
50      len =      1821      nex =      5
```

Init	32540	32883	+	0
------	-------	-------	---	---

Intr	32975	33099	+	0
------	-------	-------	---	---

Intr	33248	33493	+	0
------	-------	-------	---	---

55	Intr	33646	33901	+	0
----	------	-------	-------	---	---

Term	33987	34360	+	0
------	-------	-------	---	---

>3046853 /18389

```
60      len =      1810      nex =          5
```

1083

	Init	32589	32883	+	0
	Intr	32975	33099	+	0
	Intr	33248	33493	+	0
5	Intr	33646	33901	+	0
	Term	33987	34392	+	0
	>3046853 /21929				
10	len =	370	nex =	1	
	Sngl	34040	34389	+	0
	>3046853 /102017				
15	len =	1727	nex =	2	
	Init	63162	63584	+	0
	Term	64256	64888	+	0
	>3046853 /22470				
	len =	1319	nex =	4	
20					
25	Term	69525	69198	-	0
	Intr	70117	69922	-	0
	Intr	70247	70205	-	0
	Init	70516	70343	-	0
30	>3046854 /10299				
	len =	1773	nex =	2	
	Init	19008	19865	+	0
35	Term	20555	20780	+	0
	>3046854 /11612				
	len =	1480	nex =	3	
40					
	Init	21912	22049	+	0
	Intr	22395	22463	+	0
	Term	22582	23391	+	0
45	>3046854 /3994				
	len =	730	nex =	2	
	Term	23612	23317	-	0
50	Init	24037	23700	-	0
	>3046854 /34699				
	len =	1554	nex =	3	
55					
	Init	38733	39282	+	0
	Intr	39431	39721	+	0
	Term	39804	40286	+	0
60	>3046854 /1816				

2025-09-24 10:28:59

1084

	len =	1677	nex =	6	
5	Init	57210	57371	+	0
	Intr	57484	57534	+	0
	Intr	57762	57813	+	0
	Intr	58215	58273	+	0
	Intr	58430	58475	+	0
10	Term	58598	58886	+	0
	>3046854	/37233			
	len =	1164	nex =	3	
15	Term	78862	78599	-	0
	Intr	79238	78963	-	0
	Init	79762	79617	-	0
	>3046854	/38961			
20	len =	1774	nex =	3	
	Init	9436	9728	+	0
	Intr	9833	10113	+	0
25	Term	10198	10667	+	0
	>3046855	/108313			
	len =	657	nex =	1	
30	Sngl	23798	23142	-	0
	>3046855	/32984			
35	len =	970	nex =	1	
	Sngl	24087	23142	-	0
	>3046855	/14710			
40	len =	1037	nex =	3	
	Init	47198	47333	+	0
	Intr	47668	47769	+	0
45	Term	47850	48073	+	0
	>3046855	/143364			
	len =	697	nex =	3	
50	Init	47201	47333	+	0
	Intr	47668	47769	+	0
	Term	47850	47897	+	0
55	>3046855	/115178			
	len =	1558	nex =	5	
	Init	51391	51541	+	0
60	Intr	51756	51855	+	0

1085

Intr	52045	52191	+	0
Intr	52278	52358	+	0
Term	52709	52948	+	0

5 >3046855 /1440

```
len = 743    nex = 3
```

	Term	61068	60893	-	0
10	Intr	61285	61159	-	0
	Init	61635	61504	-	0

>3046855 /151892

```
15      len =      970      nex =      2
```

Term	61285	60730	-	0
Init	61692	61504	-	0

20 >3046855 /20686

```
len =      670    nex =      3
```

	Term	61068	61024	-	0
25	Intr	61285	61159	-	0
	Init	61692	61504	-	0

>3046855 /9633

```
30      len =      937      nex =      3
```

Term	61068	60759	-	0
Intr	61285	61159	-	0
Init	61695	61504	-	0

35
 >3046855 /107804

```
len =      911      nex =      3
```

40	Term	61068	60785	-	0
	Intr	61285	61159	-	0
	Init	61695	61504	-	0

>3046855 /21689

```
len = 1465 nex = 4
```

Init	63956	64174	+	(
Intr	64250	64483	+	(
Intr	64572	64658	+	(
Term	64758	65420	+	(

>3046855 /92780

```
55      len =      810      nex =      2
```

Init	64590	64658	+	(
Term	64758	65399	+	(

60 >3046855 /116709

1086

	len =	776	nex =	3		
5	Init	72683	72877	+	0	
	Intr	72984	73124	+	0	
	Term	73257	73458	+	0	
	>3046855		/36624			
10	len =	3759	nex =	4		
	Term	74593	73489	-	0	
	Intr	75227	74671	-	0	
	Intr	76973	76756	-	0	
15	Init	77247	77060	-	0	
	>3046856		/19567			
20	len =	1461	nex =	3		
	Init	10176	10234	+	0	
	Intr	10433	11125	+	0	
	Term	11211	11636	+	0	
25	>3046856		/15527			
	len =	550	nex =	1		
30	Sngl	20236	19688	-	0	
	>3046856		/103939			
	len =	1469	nex =	1		
35	Sngl	41267	41096	-	0	
	>3046856		/21741			
40	len =	1522	nex =	5		
	Term	40068	39764	-	0	
	Intr	40233	40144	-	0	
	Intr	40498	40331	-	0	
	Intr	40851	40586	-	0	
45	Init	41285	41096	-	0	
	>3046856		/39378			
50	len =	2380	nex =	7		
	Init	42386	42447	+	0	
	Intr	42533	42663	+	0	
	Intr	42914	42997	+	0	
	Intr	43341	43451	+	0	
55	Intr	43540	43731	+	0	
	Intr	43935	44032	+	0	
	Term	44339	44765	+	0	
60	>3046856		/25234			

1087

	len =	839	nex =	2	
	Init	43935	44032	+	0
5	Term	44339	44773	+	0
	>3046856	/7861			
	len =	2553	nex =	3	
10	Term	61214	60794	-	0
	Intr	61683	61434	-	0
	Init	63346	61947	-	0
	>3046856	/34558			
15	len =	2860	nex =	9	
	Term	65245	64782	-	0
	Intr	65449	65358	-	0
20	Intr	65643	65616	-	0
	Intr	65778	65739	-	0
	Intr	65936	65875	-	0
	Intr	66094	66031	-	0
25	Intr	66219	66181	-	0
	Intr	66631	66577	-	0
	Init	66928	66836	-	0
	>3046856	/33481			
30	len =	1393	nex =	3	
	Term	74793	73895	-	0
	Intr	75040	74878	-	0
	Init	75287	75126	-	0
35	>3047060	/107700			
	len =	1095	nex =	2	
40	Init	27099	27748	+	0
	Term	27822	28193	+	0
	>3047060	/42538			
45	len =	675	nex =	1	
	Sngl	30075	29401	-	0
	>3047074	/30174			
50	len =	518	nex =	1	
	Sngl	102999	102484	-	0
55	>3047074	/3550			
	len =	2660	nex =	6	
	Term	104534	103979	-	0
60	Intr	104851	104618	-	0

1088

	Intr	105248	104963	-	0
	Intr	105650	105501	-	0
	Intr	105885	105812	-	0
	Init	106638	106431	-	0
5	>3047074 /21874				
	len =	2304	nex =	12	
10	Term	77864	77757	-	0
	Intr	78105	77960	-	0
	Intr	78294	78189	-	0
	Intr	78454	78377	-	0
	Intr	78603	78550	-	0
15	Intr	78757	78697	-	0
	Intr	78979	78855	-	0
	Intr	79155	79069	-	0
	Intr	79339	79266	-	0
	Intr	79549	79420	-	0
20	Intr	79716	79634	-	0
	Init	80060	79919	-	0
	>3047074 /14379				
25	len =	1310	nex =	2	
	Init	88948	89130	+	0
	Term	89599	90257	+	0
30	>3047074 /33112				
	len =	596	nex =	1	
	Sngl	89686	90281	+	0
35	>3047088 /38281				
	len =	1605	nex =	4	
40	Init	13957	14138	+	0
	Intr	14229	14418	+	0
	Intr	14734	14808	+	0
	Term	14903	15561	+	0
45	>3047088 /40501				
	len =	2311	nex =	4	
	Term	16023	15615	-	0
50	Intr	16385	16228	-	0
	Intr	16857	16804	-	0
	Init	17417	16963	-	0
	>3047088 /20286				
55	len =	4330	nex =	16	
	Init	76297	76485	+	0
	Intr	76590	76926	+	0
60	Intr	77004	77145	+	0

ORIGINAL RELEASED

1089

	Intr	77287	77385	+	0
	Intr	77477	77533	+	0
	Intr	77737	77808	+	0
	Intr	78172	78261	+	0
5	Intr	78398	78457	+	0
	Intr	78763	78948	+	0
	Intr	79033	79125	+	0
	Intr	79204	79285	+	0
	Intr	79374	79480	+	0
10	Intr	79696	79771	+	0
	Intr	79926	79972	+	0
	Intr	80061	80123	+	0
	Term	80359	80623	+	0

15 >3047088 /33462

len = 1594 nex = 7

	Init	79033	79125	+	0
20	Intr	79204	79285	+	0
	Intr	79374	79480	+	0
	Intr	79696	79771	+	0
	Intr	79926	79972	+	0
25	Intr	80061	80123	+	0
	Term	80359	80623	+	0

>3047100 /34671

len = 3995 nex = 16

30	Init	15513	15554	+	0
	Intr	15672	15764	+	0
	Intr	15953	16020	+	0
	Intr	16133	16208	+	0
35	Intr	16307	16353	+	0
	Intr	16466	16582	+	0
	Intr	16799	16876	+	0
	Intr	17024	17111	+	0
	Intr	17262	17374	+	0
40	Intr	17513	17575	+	0
	Intr	17785	17889	+	0
	Intr	17986	18045	+	0
	Intr	18316	18417	+	0
45	Intr	18618	18666	+	0
	Intr	19056	19124	+	0
	Term	19230	19507	+	0

>3047100 /9233

50 len = 1124 nex = 2

	Init	19593	19904	+	0
	Term	20386	20716	+	0

55 >3047100 /31467

len = 730 nex = 2

60	Term	20944	20738	-	0
	Init	21458	21292	-	0

2025-07-25 10:00:00

1090

>3047100 /18459

5	len =	1570	nex =	1	
	Sngl	39059	38782	-	0

>3047100 /18629

10	len =	1636	nex =	1	
	Sngl	39059	38776	-	0

>3047100 /545

15	len =	498	nex =	2	
	Term	67746	67593	-	0
	Init	68090	67824	-	0

20 >3047100 /26070

	len =	1097	nex =	3	
25	Term	67746	67416	-	0
	Intr	68089	67824	-	0
	Init	68512	68419	-	0

>3047100 /28642

30	len =	3207	nex =	9	
	Init	78779	78999	+	0
	Intr	79083	79218	+	0
35	Intr	79317	79450	+	0
	Intr	79542	79661	+	0
	Intr	79739	79824	+	0
	Intr	79921	80109	+	0
	Intr	80200	80352	+	0
40	Intr	80464	80577	+	0
	Term	80658	81268	+	0

>3056579 /158942

45	len =	831	nex =	1	
	Sngl	32993	33823	+	0

>3056579 /13461

50	len =	3450	nex =	7	
	Init	42933	44316	+	0
	Intr	44426	44536	+	0
55	Intr	44764	44900	+	0
	Intr	44977	45187	+	0
	Intr	45292	45529	+	0
	Intr	45624	45774	+	0
60	Term	45872	46382	+	0

2025-09-24 10:00:00

>3056579 /38645

len = 1408 nex = 3

5	Term	62928	62777	-	0
	Intr	63094	63020	-	0
	Init	63938	63662	-	0

>3059018 /29133

10 len = 1298 nex = 2

	Term	19437	19300	-	0
	Init	19744	19558	-	0

15 >3059018 /20592

len = 2679 nex = 4

20	Init	70884	70974	+	0
	Intr	72180	72269	+	0
	Intr	72361	72445	+	0
	Term	72984	73562	+	0

25 >3059018 /38430

len = 1379 nex = 1

	Sngl	72180	72270	+	0
--	------	-------	-------	---	---

30 >3059018 /38689

len = 610 nex = 1

35 Sngl 82744 83198 + 0

>3059018 /18947

len = 1732 nex = 3

40	Init	87586	88234	+	0
	Intr	88340	88417	+	0
	Term	88957	89317	+	0

45 >3063438 /7188

len = 861 nex = 1

	Sngl	106806	107666	+	0
--	------	--------	--------	---	---

50 >3063438 /32337

len = 1841 nex = 4

55	Init	10816	11218	+	0
	Intr	11470	11646	+	0
	Intr	11732	11896	+	0
	Term	11987	12656	+	0

60 >3063438 /20276

2025 RELEASE UNDER E.O. 14176

1092

	len =	1412	nex =	3	
5	Term	108108	107703	-	0
	Intr	108949	108833	-	0
	Init	109114	109038	-	0
	>3063438 /33493				
10	len =	2171	nex =	7	
	Init	125603	125698	+	0
	Intr	125795	125896	+	0
	Intr	126029	126124	+	0
15	Intr	126303	126334	+	0
	Intr	126429	126504	+	0
	Intr	126676	126729	+	0
	Term	126954	127179	+	0
20	>3063438 /118748				
	len =	2133	nex =	7	
	Init	125603	125698	+	0
25	Intr	125795	125896	+	0
	Intr	126029	126124	+	0
	Intr	126303	126334	+	0
	Intr	126429	126504	+	0
	Intr	126676	126729	+	0
30	Term	126954	127173	+	0
	>3063438 /36523				
	len =	2136	nex =	7	
35	Init	125603	125698	+	0
	Intr	125795	125896	+	0
	Intr	126029	126124	+	0
	Intr	126303	126334	+	0
40	Intr	126429	126504	+	0
	Intr	126676	126729	+	0
	Term	126954	127176	+	0
	>3063438 /34741				
45	len =	1210	nex =	2	
	Term	13932	13866	-	0
	Init	14277	14113	-	0
50	>3063438 /35718				
	len =	2841	nex =	9	
55	Term	47185	47114	-	0
	Intr	47500	47405	-	0
	Intr	47724	47599	-	0
	Intr	47999	47942	-	0
	Intr	48148	48093	-	0
60	Intr	48315	48238	-	0

2025 RELEASE UNDER E.O. 14176

1093

Intr	48516	48412	-	0
Intr	48701	48602	-	0
Init	49201	49002	-	0

5	>3063438	/43026		
	len =	1060	nex =	1
10	Sngl	68860	67801	- 0
	>3063438	/27805		
	len =	1272	nex =	1
15	Sngl	8642	7563	- 0
	>3063690	/40949		
	len =	2193	nex =	3
20	Init	18843	19345	+ 0
	Intr	19886	19966	+ 0
	Term	20222	21035	+ 0
25	>3063690	/18482		
	len =	3096	nex =	5
	Term	21537	21211	- 0
30	Intr	21947	21729	- 0
	Intr	22174	22060	- 0
	Intr	22529	22282	- 0
	Init	24306	22854	- 0
35	>3063690	/35221		
	len =	1829	nex =	2
	Init	51969	52232	+ 0
40	Term	52962	53797	+ 0
	>3063690	/39535		
	len =	1673	nex =	3
45	Init	89115	89369	+ 0
	Intr	89531	89856	+ 0
	Term	90448	90787	+ 0
50	>3063690	/28609		
	len =	1510	nex =	3
	Term	94514	94452	- 0
55	Intr	95366	95253	- 0
	Init	95700	95452	- 0
	>3068702	/21630		
60	len =	1352	nex =	4

2025 RELEASE UNDER E.O. 14176

1094

	Term	9605	9217	-	0
	Intr	9840	9694	-	0
5	Intr	10174	9918	-	0
	Init	10568	10350	-	0
>3068702		/38567			
	len =	1554	nex =	4	
10	Term	9605	9176	-	0
	Intr	9840	9694	-	0
	Intr	10174	9918	-	0
	Init	10729	10350	-	0
15	>3068702	/6926			
	len =	872	nex =	2	
20	Term	12518	12287	-	0
	Init	13158	12549	-	0
>3068702		/3844			
25	len =	3670	nex =	12	
	Term	28154	27776	-	0
	Intr	28332	28240	-	0
30	Intr	28985	28814	-	0
	Intr	29153	29077	-	0
	Intr	29837	29728	-	0
	Intr	30000	29941	-	0
	Intr	30183	30110	-	0
35	Intr	30416	30293	-	0
	Intr	30579	30502	-	0
	Intr	30782	30676	-	0
	Intr	30970	30866	-	0
	Init	31445	31360	-	0
40	>3068702	/93427			
	len =	1415	nex =	3	
45	Term	52373	51897	-	0
	Intr	52618	52465	-	0
	Init	53311	52956	-	0
>3075383		/37427			
50	len =	2556	nex =	3	
	Init	47142	47616	+	0
	Intr	48151	48776	+	0
	Term	48851	49697	+	0
55	>3075383	/4752			
	len =	562	nex =	1	
60	Snq1	49196	49757	+	0

1095

>3075383 /33535

5	len =	1150	nex =	3	
	Init	51795	51949	+	0
	Intr	52409	52543	+	0
	Term	52636	52942	+	0

10 >3075383 /38538

	len =	2558	nex =	8	
	Term	53383	53089	-	0
15	Intr	53656	53563	-	0
	Intr	53916	53743	-	0
	Intr	54085	53994	-	0
	Intr	54632	54523	-	0
	Intr	54787	54712	-	0
20	Intr	54992	54900	-	0
	Init	55322	55106	-	0

>3080352 /18098

25	len =	1590	nex =	8	
	Term	40284	40098	-	0
	Intr	40461	40369	-	0
	Intr	40639	40558	-	0
30	Intr	40838	40768	-	0
	Intr	40960	40928	-	0
	Intr	41121	41092	-	0
	Intr	41295	41227	-	0
35	Init	41687	41583	-	0

>3080352 /38902

	len =	1694	nex =	8	
40	Term	40284	40037	-	0
	Intr	40461	40369	-	0
	Intr	40639	40558	-	0
	Intr	40838	40768	-	0
	Intr	40960	40928	-	0
45	Intr	41121	41092	-	0
	Intr	41295	41227	-	0
	Init	41730	41583	-	0

>3080352 /39492

50	len =	1192	nex =	1	
	Sngl	56521	57712	+	0

55 >3080352 /119432

	len =	1076	nex =	4	
	Term	59135	58928	-	0
60	Intr	59332	59218	-	0

2025 RELEASE UNDER E.O. 14176

1096

Intr	59524	59415	-	0
Init	60003	59857	-	0

>3080352 /34674

5

len = 3652 nex = 12

Term	60315	60098	-	0
Intr	60539	60394	-	0
Intr	60747	60642	-	0
Intr	60927	60850	-	0
Intr	61080	61027	-	0
Intr	61235	61175	-	0
Intr	61529	61405	-	0
Intr	61701	61615	-	0
Intr	61918	61839	-	0
Intr	62170	62041	-	0
Intr	62546	62464	-	0
Init	62781	62669	-	0

10

15

20

>3080352 /11264

len = 2126 nex = 10

25

Term	65646	65231	-	0
Intr	65897	65752	-	0
Intr	66105	66000	-	0
Intr	66302	66225	-	0
Intr	66445	66392	-	0
Intr	66599	66539	-	0
Intr	66819	66695	-	0
Intr	67006	66920	-	0
Intr	67165	67086	-	0
Init	67356	67258	-	0

35

>3080352 /36325

len = 3216 nex = 12

40

Term	65646	65229	-	0
Intr	65897	65752	-	0
Intr	66105	66000	-	0
Intr	66302	66225	-	0
Intr	66445	66392	-	0
Intr	66599	66539	-	0
Intr	66819	66695	-	0
Intr	67006	66920	-	0
Intr	67165	67086	-	0
Intr	67387	67258	-	0
Intr	67607	67525	-	0
Init	67822	67707	-	0

50

>3080352 /13917

55 len = 681 nex = 1

Sngl 77159 76479 - 0

>3080406 /32793

60

1097

	len =	1690	nex =	5	
	Init	3323	3465	+	0
	Intr	3592	3780	+	0
5	Intr	3945	4107	+	0
	Intr	4211	4403	+	0
	Term	4497	4800	+	0
	>3080406 /37017				
10	len =	2965	nex =	11	
	Term	28768	28521	-	0
	Intr	29006	28912	-	0
15	Intr	29136	29088	-	0
	Intr	29290	29231	-	0
	Intr	29583	29462	-	0
	Intr	29930	29882	-	0
	Intr	30171	30100	-	0
20	Intr	30408	30308	-	0
	Intr	31062	30951	-	0
	Intr	31203	31134	-	0
	Init	31485	31423	-	0
25	>3080406 /22906				
	len =	1493	nex =	5	
	Term	5013	4742	-	0
30	Intr	5195	5109	-	0
	Intr	5342	5307	-	0
	Intr	5651	5451	-	0
	Init	5811	5725	-	0
35	>3080406 /12228				
	len =	2116	nex =	4	
	Init	62410	63011	+	0
40	Intr	63132	63371	+	0
	Intr	63788	63866	+	0
	Term	63967	64525	+	0
	>3080406 /2912				
45	len =	2158	nex =	7	
	Init	65401	65852	+	0
	Intr	65941	66039	+	0
50	Intr	66212	66406	+	0
	Intr	66528	66620	+	0
	Intr	66715	66810	+	0
	Intr	66903	67043	+	0
	Term	67158	67558	+	0
55	>3080406 /1368				
	len =	1665	nex =	6	
60	Term	84894	84458	-	0

000000-000000

1099

000100-000000

	>3080430	/17227		
5	len =	134	nex =	1
	Sngl	9916	10049	+ 0
	>3108024	/3434		
10	len =	2352	nex =	2
	Term	29167	28743	- 0
	Init	30389	29273	- 0
15	>3108024	/124122		
	len =	1576	nex =	7
20	Term	43805	43635	- 0
	Intr	43963	43896	- 0
	Intr	44150	44048	- 0
	Intr	44473	44243	- 0
	Intr	44946	44906	- 0
25	Intr	45067	45031	- 0
	Init	45200	45177	- 0
	>3108024	/2164		
30	len =	1615	nex =	7
	Term	43805	43610	- 0
	Intr	43963	43896	- 0
	Intr	44150	44048	- 0
	Intr	44473	44243	- 0
35	Intr	44946	44906	- 0
	Intr	45067	45031	- 0
	Init	45224	45177	- 0
	>3108025	/34936		
40	len =	2501	nex =	4
	Init	101875	102049	+ 0
	Intr	102345	102824	+ 0
45	Intr	102969	103448	+ 0
	Term	103539	104375	+ 0
	>3108025	/113281		
50	len =	1612	nex =	3
	Term	104623	104358	- 0
	Intr	105265	104996	- 0
	Init	105449	105341	- 0
55	>3128134	/33790		
	len =	2267	nex =	4
60	Init	16068	16558	+ 0

1100

Intr	17365	17675	+	0
Intr	17756	17820	+	0
Term	17936	18334	+	0

5 >3128134 /37644

len = 2174 nex = 6

10	Term	19620	18952	-	0
	Intr	19787	19706	-	0
	Intr	20141	20058	-	0
	Intr	20421	20245	-	0
	Intr	20669	20507	-	0
	Init	21125	20761	-	0

15 >3128134 /42970

len = 1721 nex = 2

20	Init	22577	22849	+	0
	Term	22987	24297	+	0

>3128134 /39030

25	len =	619	nex =	1	
	Sngl	6536	7154	+	0

>3128135 /33791

30	len =	1822	nex =	2	
	Init	21696	22080	+	0
	Term	22757	23517	+	0

35 >3128135 /25162

len = 1581 nex = 4

40	Term	47134	46795	-	0
	Intr	47507	47403	-	0
	Intr	47751	47624	-	0
	Init	48188	48109	-	0

45 >3128135 /39130

len = 1611 nex = 5

50	Term	47134	46806	-	0
	Intr	47507	47403	-	0
	Intr	47751	47624	-	0
	Intr	48188	48109	-	0
	Init	48416	48288	-	0

55 >3128135 /18718

len = 1690 nex = 3

60	Init	48812	49068	+	0
	Intr	49281	49599	+	0

1101

	Term	50104	50162	+	0
	>3128135 /110653				
5	len =	1658	nex =	5	
	Init	53406	53626	+	0
	Intr	53710	53762	+	0
	Intr	53845	54012	+	0
10	Intr	54431	54538	+	0
	Term	54771	55063	+	0
	>3128135 /1455				
15	len =	2771	nex =	11	
	Term	55477	55365	-	0
	Intr	55672	55601	-	0
	Intr	55818	55768	-	0
20	Intr	56007	55920	-	0
	Intr	56234	56111	-	0
	Intr	56410	56338	-	0
	Intr	56574	56511	-	0
	Intr	56855	56662	-	0
25	Intr	56992	56943	-	0
	Intr	57191	57128	-	0
	Init	57772	57284	-	0
	>3128135 /6216				
30	len =	3834	nex =	15	
	Init	58058	58788	+	0
	Intr	58941	59002	+	0
35	Intr	59123	59223	+	0
	Intr	59534	59584	+	0
	Intr	59679	59753	+	0
	Intr	59861	59938	+	0
	Intr	60097	60153	+	0
40	Intr	60287	60338	+	0
	Intr	60449	60519	+	0
	Intr	60620	60719	+	0
	Intr	60816	60920	+	0
	Intr	61011	61073	+	0
45	Intr	61154	61276	+	0
	Intr	61370	61531	+	0
	Term	61621	61891	+	0
	>3128135 /235				
50	len =	1547	nex =	3	
	Init	62091	62463	+	0
	Intr	62667	63009	+	0
55	Term	63093	63637	+	0
	>3128136 /35906				
60	len =	738	nex =	1	

1102

Sngl	1	738	+	0
------	---	-----	---	---

>3128136 /3080

```
5      len = 1342      nex = 4
```

Init	23530	24049	+	0
------	-------	-------	---	---

Intr	24161	24386	+	0
------	-------	-------	---	---

Intr	24474	24547	+	0
------	-------	-------	---	---

10	Term	24640	24871	+	0
----	------	-------	-------	---	---

>3128136 /2222

```
len = 1300    nex = 4
```

Init	23572	24049	+	0
------	-------	-------	---	---

Intr	24161	24386	+	0
------	-------	-------	---	---

Intr	24474	24547	+	0
------	-------	-------	---	---

Term	24640	24871	+	0
------	-------	-------	---	---

>3128136 /38030

```
len = 2477    nex = 5
```

25	Term	28271	27313	-	0
----	------	-------	-------	---	---

Intr	28564	28362	-	0
------	-------	-------	---	---

Intr	28798	28676	-	0
------	-------	-------	---	---

Intr	29081	28887	-	0
------	-------	-------	---	---

Init	29789	29424	-	0
------	-------	-------	---	---

>3128136 /93374

```
len =      550    nex =      1
```

35	Sngl	36684	36143	-	0
----	------	-------	-------	---	---

>3128136 /34700

```
len = 1901 nex = 5
```

Term	39171	38646	-	(
------	-------	-------	---	---

Intr	39414	39263	-	(
------	-------	-------	---	---

Intr	39628	39503	-	(
------	-------	-------	---	---

Intr	39879	39736	-	(
------	-------	-------	---	---

Init	40546	40298	-	
------	-------	-------	---	--

>3128136 /35310

len = 1990 nex = 9

Term	42802	42663	-
------	-------	-------	---

Intr	42973	42892	-
------	-------	-------	---

Intr	43171	43064	-
------	-------	-------	---

Intr	43366	43269	-
------	-------	-------	---

Intr	43569	43471	-
------	-------	-------	---

Intr	43782	43645	-
------	-------	-------	---

Intr	43925	43874	-
------	-------	-------	---

Intr	44136	44023	-
------	-------	-------	---

Init	44318	44220	-
------	-------	-------	---

1103

>3128136 /15350

len = 3109 nex = 11

5	Term	45887	45746	-	0
	Intr	46034	45984	-	0
	Intr	46201	46130	-	0
	Intr	46384	46316	-	0
	Intr	46520	46479	-	0
10	Intr	46680	46612	-	0
	Intr	46901	46764	-	0
	Intr	47104	46994	-	0
	Intr	47275	47188	-	0
	Intr	47451	47396	-	0
15	Init	48367	47549	-	0

>3128136 /37218

len = 1584 nex = 3

20	Init	57574	57902	+	0
	Intr	58199	58492	+	0
	Term	58579	59157	+	0

25 >3128136 /149202

len = 1438 nex = 2

30	Init	57626	57902	+	0
	Term	58579	59063	+	0

>3128136 /31524

len = 750 nex = 1

35	Sngl	61551	62300	+	0
----	------	-------	-------	---	---

>3128136 /2152

len = 681 nex = 1

40	Sngl	63097	63777	+	0
----	------	-------	-------	---	---

>3128136 /2489

len = 1832 nex = 7

	Init	8722	8908	+	0
	Intr	9213	9287	+	0
50	Intr	9396	9443	+	0
	Intr	9532	9688	+	0
	Intr	9777	9849	+	0
	Intr	10088	10185	+	0
	Term	10266	10553	+	0

55 >3128137 /1311

len = 670 nex = 1

60	Sngl	10381	10111	-	0
----	------	-------	-------	---	---

1104

>3128137 /2518

```

5      len =      509      nex =      1
      Snq1 13914 13406      -      0

```

>3128137 /2561

```
10      len =      639      nex =      1
        Snq1 17617 16979      -      0
```

>3128137 /35979

15	len =	1795	nex =	6	
	Init	29554	29782	+	0
	Intr	30223	30304	+	0
20	Intr	30392	30424	+	0
	Intr	30505	30626	+	0
	Intr	30713	30874	+	0
	Term	30962	31348	+	0

25 >3128137 /25262

```

len =      2145    nex =      4
30      Term 35368  34765    -    0
      Intr 35549  35466    -    0
      Intr 35707  35623    -    0
      Init 36012  35792    -    0

```

>3128137 /28455

	len =	1527	nex =	5	
	Init	37172	37397	+	0
	Intr	37545	37633	+	0
40	Intr	37818	37947	+	0
	Intr	38129	38158	+	0
	Term	38466	38698	+	0

>3128137 /9946

```
len =      719    nex =      2
Term 39259 39103 - 0
Init 39475 39347 - 0
```

50
- >3128138 /14654

	len =	953	nex =	4	
55	Term	33774	33527	-	0
	Intr	33934	33895	-	0
	Intr	34248	34172	-	0
	Init	34479	34388	-	0

60 >3128139 /6095

1105

	len =	203	nex =	1	
5	Sngl	34265	34183	-	0
	>3128139	/9938			
	len =	769	nex =	3	
10	Term	33664	33501	-	0
	Intr	34085	33948	-	0
	Init	34269	34183	-	0
	>3128139	/114411			
15	len =	835	nex =	4	
	Init	36332	36381	+	0
	Intr	36467	36596	+	0
20	Intr	36706	36764	+	0
	Term	36863	37166	+	0
	>3128139	/100570			
25	len =	408	nex =	1	
	Sngl	42429	42022	-	0
	>3128139	/30782			
30	len =	1108	nex =	0	
	>3128139	/14992			
35	len =	1135	nex =	1	
	Sngl	42487	42041	-	0
	>3128139	/117908			
40	len =	101	nex =	1	
	Sngl	53867	53767	-	0
45	>3128139	/36495			
	len =	3037	nex =	10	
	Init	56082	56250	+	0
50	Intr	57096	57195	+	0
	Intr	57270	57367	+	0
	Intr	57458	57546	+	0
	Intr	57631	57691	+	0
	Intr	57783	57947	+	0
55	Intr	58040	58096	+	0
	Intr	58194	58249	+	0
	Intr	58361	58496	+	0
	Term	58593	59118	+	0
60	>3128139	/119783			

000000-000000

2025 RELEASE UNDER E.O. 14176

1108

	len =	398	nex =	1	
	Sngl	50607	50210	-	0
5	>3128142	/31032			
	len =	3178	nex =	9	
10	Init	55780	56016	+	0
	Intr	56983	57164	+	0
	Intr	57277	57357	+	0
	Intr	57442	57501	+	0
15	Intr	57586	57663	+	0
	Intr	57752	57815	+	0
	Intr	57898	58169	+	0
	Intr	58258	58485	+	0
	Term	58560	58957	+	0
20	>3128142	/16740			
	len =	1431	nex =	2	
	Init	59192	59602	+	0
25	Term	60287	60622	+	0
	>3128142	/30437			
	len =	1315	nex =	2	
30	Init	81339	81632	+	0
	Term	82066	82653	+	0
	>3128142	/8294			
35	len =	1150	nex =	2	
	Term	84674	84011	-	0
	Init	85146	84757	-	0
40	>3128143	/39401			
	len =	3813	nex =	11	
45	Init	16973	17402	+	0
	Intr	17496	17719	+	0
	Intr	17842	18089	+	0
	Intr	18179	18398	+	0
50	Intr	18712	18839	+	0
	Intr	19099	19355	+	0
	Intr	19458	19565	+	0
	Intr	19668	19874	+	0
	Intr	19978	20088	+	0
	Intr	20176	20349	+	0
55	Term	20425	20785	+	0
	>3128143	/24169			
	len =	1371	nex =	2	
60					

1109

Term	21558	21040	-	0
Init	22410	21729	-	0

>3128143 /22388

5	len =	1134	nex =	3	
	Init	23440	23527	+	0
	Intr	23989	24077	+	0
10	Term	24347	24573	+	0

>3128143 /97304

15	len =	1618	nex =	4	
	Init	34719	34756	+	0
	Intr	35037	35222	+	0
	Intr	35369	35408	+	0
20	Term	35792	36036	+	0

>3128143 /17995

	len =	2075	nex =	3	
25	Init	38456	38783	+	0
	Intr	38871	39113	+	0
	Term	39198	40530	+	0

>3128143 /5688

30	len =	518	nex =	1	
	Sngl	43484	44001	+	0

>3128143 /41898

	len =	416	nex =	1	
40	Sngl	43585	44000	+	0

>3128143 /11837

	len =	2078	nex =	8	
45	Term	44515	44311	-	0
	Intr	44724	44603	-	0
	Intr	44933	44823	-	0
	Intr	45130	45020	-	0
	Intr	45250	45213	-	0
50	Intr	45393	45323	-	0
	Intr	46203	46049	-	0
	Init	46388	46280	-	0

>3128143 /6495

55	len =	2273	nex =	8	
	Term	44515	44170	-	0
	Intr	44724	44603	-	0
60	Intr	44933	44823	-	0

1110

	Intr	45130	45020	-	0
	Intr	45250	45213	-	0
	Intr	45393	45323	-	0
	Intr	46203	46049	-	0
5	Init	46442	46280	-	0
	>3128143		/43069		
10	len =	1223	nex =	2	
	Term	47252	46474	-	0
	Init	47696	47350	-	0
15	>3128143		/31005		
	len =	1417	nex =	4	
	Term	48228	47939	-	0
	Intr	48546	48312	-	0
20	Intr	48718	48648	-	0
	Init	49355	49116	-	0
	>3128143		/4502		
25	len =	1283	nex =	1	
	Sngl	54087	54642	+	0
30	>3128143		/3836		
	len =	1672	nex =	5	
	Init	57472	57668	+	0
	Intr	57754	57814	+	0
35	Intr	58106	58210	+	0
	Intr	58389	58672	+	0
	Term	58870	59134	+	0
40	>3128143		/5915		
	len =	563	nex =	3	
	Init	78978	79080	+	0
	Intr	79169	79306	+	0
45	Term	79392	79540	+	0
	>3128166		/34876		
50	len =	2304	nex =	7	
	Init	104985	105384	+	0
	Intr	105472	105604	+	0
	Intr	105754	105891	+	0
	Intr	105992	106125	+	0
55	Intr	106247	106372	+	0
	Intr	106603	106706	+	0
	Term	107114	107288	+	0
60	>3128166		/8402		

1111

	len =	1414	nex =	5	
	Init	107501	107707	+	0
	Intr	107864	107980	+	0
5	Intr	108265	108299	+	0
	Intr	108441	108541	+	0
	Term	108639	108914	+	0
>3128166 /9014					
10	len =	2503	nex =	8	
	Init	109037	109232	+	0
	Intr	109367	109543	+	0
15	Intr	109771	109860	+	0
	Intr	110290	110433	+	0
	Intr	110541	110708	+	0
	Intr	110780	110849	+	0
20	Intr	110930	111011	+	0
	Term	111101	111260	+	0
>3128166 /91801					
25	len =	655	nex =	2	
	Term	23650	23349	-	0
	Init	23856	23745	-	0
>3128166 /18124					
30	len =	2278	nex =	10	
	Term	23650	23314	-	0
	Intr	23856	23745	-	0
35	Intr	24003	23937	-	0
	Intr	24158	24112	-	0
	Intr	24328	24263	-	0
	Intr	24527	24415	-	0
40	Intr	24742	24623	-	0
	Intr	24972	24830	-	0
	Intr	25156	25042	-	0
	Init	25591	25463	-	0
>3128166 /2895					
45	len =	705	nex =	2	
	Term	31097	30635	-	0
	Init	31334	31265	-	0
50	>3128166 /26701				
	len =	599	nex =	1	
55	Sngl	34617	35215	+	0
>3128166 /16528					
60	len =	4456	nex =	19	

2025 RELEASE UNDER E.O. 14176

1112

5 Init 42061 42128 + 0
 Intr 42230 42292 + 0
 Intr 42439 42480 + 0
 Intr 42662 42727 + 0
 Intr 42820 42885 + 0
 Intr 43008 43121 + 0
 Intr 43230 43293 + 0
 Intr 43471 43553 + 0
 10 Intr 43682 43731 + 0
 Intr 43811 43888 + 0
 Intr 44033 44161 + 0
 Intr 44296 44338 + 0
 Intr 44430 44516 + 0
 Intr 44684 44785 + 0
 15 Intr 45134 45241 + 0
 Intr 45333 45383 + 0
 Intr 45531 45604 + 0
 Intr 45682 45733 + 0
 Term 45831 46160 + 0
 20 >3128166 /40257
 len = 474 nex = 1
 25 Sngl 47811 47358 - 0
 >3128166 /23768
 len = 970 nex = 0
 30 >3128166 /21043
 len = 2437 nex = 9
 35 Init 55878 56041 + 0
 Intr 56251 56433 + 0
 Intr 56538 56602 + 0
 Intr 56676 56775 + 0
 Intr 56882 56920 + 0
 40 Intr 57009 57191 + 0
 Intr 57460 57638 + 0
 Intr 57783 57902 + 0
 Term 57996 58314 + 0
 45 >3128166 /39057
 len = 1992 nex = 7
 Init 6113 6363 + 0
 50 Intr 6454 6540 + 0
 Intr 6632 6895 + 0
 Intr 6985 7149 + 0
 Intr 7228 7281 + 0
 Intr 7374 7520 + 0
 55 Term 7616 8104 + 0
 >3132469 /14874
 len = 2445 nex = 6
 60

1113

	Term	4589	4297	-	0
	Intr	4838	4692	-	0
	Intr	5021	4916	-	0
	Intr	5351	5155	-	0
5	Intr	5552	5460	-	0
	Init	6741	5705	-	0
>3133272		/15372			
10	len =	596	nex =	1	
	Sngl	15821	16416	+	0
>3133272		/105334			
15	len =	190	nex =	1	
	Sngl	47085	47265	+	0
>3135250		/99763			
20	len =	598	nex =	1	
	Sngl	34775	34178	-	0
25	>3135250		/37787		
	len =	1140	nex =	2	
30	Init	40584	40900	+	0
	Term	41358	41723	+	0
>3150395		/3754			
35	len =	708	nex =	1	
	Sngl	12833	13540	+	0
>3150395		/37527			
40	len =	1938	nex =	5	
	Term	19470	19016	-	0
	Intr	19706	19545	-	0
45	Intr	20319	19790	-	0
	Intr	20522	20409	-	0
	Init	20953	20632	-	0
>3150395		/34995			
50	len =	2070	nex =	5	
	Term	21605	21212	-	0
	Intr	21821	21733	-	0
55	Intr	22028	21935	-	0
	Intr	22801	22629	-	0
	Init	23281	23141	-	0
>3150395		/106011			
60					

1114

	len =	2110	nex =	5	
	Term	21605	21183	-	0
	Intr	21821	21733	-	0
5	Intr	22028	21935	-	0
	Intr	22801	22629	-	0
	Init	23286	23141	-	0
	>3150395 /22034				
10	len =	1351	nex =	3	
	Term	23931	23690	-	0
	Intr	24424	24224	-	0
15	Init	25040	24615	-	0
	>3150395 /1853				
	len =	1297	nex =	3	
20	Init	75585	75743	+	0
	Intr	75875	75961	+	0
	Term	76230	76442	+	0
25	>3150396 /299				
	len =	2574	nex =	10	
	Init	20107	20448	+	0
30	Intr	20697	20782	+	0
	Intr	20860	20926	+	0
	Intr	21087	21229	+	0
	Intr	21397	21445	+	0
	Intr	21591	21728	+	0
35	Intr	21813	21893	+	0
	Intr	21979	22155	+	0
	Intr	22242	22328	+	0
	Term	22416	22680	+	0
40	>3150396 /37777				
	len =	1499	nex =	7	
	Init	21157	21229	+	0
45	Intr	21397	21445	+	0
	Intr	21591	21728	+	0
	Intr	21813	21893	+	0
	Intr	21979	22155	+	0
	Intr	22242	22328	+	0
50	Term	22416	22655	+	0
	>3150396 /18023				
	len =	550	nex =	1	
55	Sngl	29680	30224	+	0
	>3150396 /29970				
60	len =	1095	nex =	2	

1115

	Term	36375	35916	-	0
	Init	37010	36739	-	0
5	>3150396	/4576			
	len =	1543	nex =	2	
10	Init	65489	66226	+	0
	Term	66584	67031	+	0
	>3150396	/29581			
15	len =	1572	nex =	4	
	Term	77303	77124	-	0
	Intr	78071	77946	-	0
	Intr	78321	78170	-	0
	Init	78695	78445	-	0
20	>3150396	/36577			
	len =	2177	nex =	1	
25	Sngl	80119	79013	-	0
	>3152602	/33637			
30	len =	1336	nex =	2	
	Init	12908	13530	+	0
	Term	13557	14243	+	0
35	>3152602	/17402			
	len =	1296	nex =	1	
	Sngl	16191	17124	+	0
40	>3152602	/2328			
	len =	651	nex =	1	
45	Sngl	68201	67551	-	0
	>3152602	/36311			
	len =	1291	nex =	1	
50	Sngl	7655	8945	+	0
	>3152602	/20125			
55	len =	690	nex =	1	
	Sngl	76571	77260	+	0
	>3152602	/104289			
60	len =	760	nex =	1	

2025 RELEASE UNDER E.O. 14176

1116

	Sngl	77189	76430	-	0
	>3169169 /41424				
5	len =	4293	nex =	10	
	Init	10438	10717	+	0
	Intr	10800	10945	+	0
10	Intr	11012	11088	+	0
	Intr	11168	11290	+	0
	Intr	11602	12013	+	0
	Intr	12378	12464	+	0
	Intr	12552	12620	+	0
15	Intr	12727	12798	+	0
	Intr	14190	14303	+	0
	Term	14407	14730	+	0
	>3169169 /12708				
20	len =	4175	nex =	12	
	Init	10565	10717	+	0
	Intr	10800	10945	+	0
25	Intr	11012	11088	+	0
	Intr	11168	11290	+	0
	Intr	11602	11704	+	0
	Intr	11912	12013	+	0
	Intr	12378	12464	+	0
30	Intr	12552	12620	+	0
	Intr	12727	12798	+	0
	Intr	13609	13730	+	0
	Intr	14190	14303	+	0
	Term	14559	14739	+	0
35	>3169169 /33165				
	len =	3649	nex =	14	
40	Init	23773	23969	+	0
	Intr	24056	24201	+	0
	Intr	24279	24355	+	0
	Intr	24441	24560	+	0
	Intr	24831	24933	+	0
45	Intr	25023	25060	+	0
	Intr	25156	25257	+	0
	Intr	25439	25525	+	0
	Intr	25617	25700	+	0
	Intr	25899	25970	+	0
50	Intr	26085	26203	+	0
	Intr	26875	26988	+	0
	Intr	27084	27133	+	0
	Term	27222	27421	+	0
55	>3169169 /38084				
	len =	2296	nex =	10	
	Init	58707	58874	+	0
60	Intr	59120	59249	+	0

1117

5	Intr	59331	59525	+	0
	Intr	59614	59771	+	0
	Intr	59925	60015	+	0
	Intr	60148	60213	+	0
	Intr	60303	60385	+	0
	Intr	60504	60567	+	0
	Intr	60639	60707	+	0
	Term	60845	61002	+	0
10	>3169169	/24000			
	len =	850	nex =	3	
15	Term	61337	61068	-	0
	Intr	61652	61612	-	0
	Init	61908	61760	-	0
	>3169169	/14468			
20	len =	953	nex =	3	
	Term	61337	61019	-	0
	Intr	61652	61612	-	0
25	Init	61971	61760	-	0
	>3169169	/27437			
	len =	407	nex =	1	
30	Sngl	73602	74008	+	0
	>3169169	/206407			
35	len =	478	nex =	1	
	Sngl	73603	74080	+	0
	>3169169	/9937			
40	len =	478	nex =	1	
	Sngl	73604	74081	+	0
	>3169169	/28423			
45	len =	507	nex =	1	
	Sngl	73604	74110	+	0
	>3169169	/26935			
50	len =	530	nex =	1	
	Sngl	73604	74133	+	0
	>3169169	/11891			
55	len =	532	nex =	1	
	Sngl	73604	74135	+	0
	>3169169	/11891			
60	len =	532	nex =	1	
	Sngl	73604	74135	+	0
	>3169169	/11891			

1118

>3169169 /121073

```

5      len =      478      nex =      1
      Snq1 73656 74133      +      0

```

>3172156 /26867

```
10      len =      790      nex =      1
      Snq1 3280      4062      +      0
```

>3172156 /7894

	len =	2323	nex =	4	
20	Init	45242	46579	+	0
	Intr	46725	46788	+	0
	Intr	47076	47119	+	0
	Term	47209	47564	+	

>3172156 /39526

25	len =	1772	nex =	4	
	Init	45807	46579	+	0
	Intr	46725	46788	+	0
	Intr	47076	47119	+	0
30	Term	47209	47578	+	0

>3172156 /6711

35	len =	1810	nex =	5	
	Init	48322	48537	+	0
	Intr	48765	48805	+	0
	Intr	48909	49207	+	0
	Intr	49460	49606	+	0
40	Term	49686	50127	+	0

>3172156 /120133

45	len =	1781	nex =	5	
	Init	48347	48537	+	0
	Intr	48765	48805	+	0
	Intr	48909	49207	+	0
	Intr	49460	49606	+	0
50	Term	49686	50127	+	0

>3172156 /114146

```

55         len =      2516      nex =      0
        >3172156      /6399

```

```

len =      2508    nex =          3
60      Init 63063    63494      +      0

```


1119

Intr	63615	64173	+	0
Term	64493	64821	+	0

Term	64493	64821	+	0
------	-------	-------	---	---

>3172156 /37537

```
5      len =      817      nex =      4
```

Term	74213	74001	-	0
------	-------	-------	---	---

Intr	74366	74320	-	0
------	-------	-------	---	---

10	Intr	74576	74517	-	0
----	------	-------	-------	---	---

Init	74817	74669	-	0
------	-------	-------	---	---

>3172156 /25408

```
15      len =      957      nex =      4
```

Term	74213	74000	-	0
------	-------	-------	---	---

Intr	74366	74320	-	0
------	-------	-------	---	---

Intr	74576	74517	-	0
------	-------	-------	---	---

Init	74956	74669	-	0
------	-------	-------	---	---

>3172156 /94230

```
len = 2068    nex = 2
```

Term	7070	6731	-	0
------	------	------	---	---

Term	875	8751		
Init	7680	7442	-	

>3176693 /145854

```
len = 1977 nex = 2
```

Init	20568	20774	+
------	-------	-------	---

Term	20878	21151	+
------	-------	-------	---

35

178833 178832

[illegible]

Init	31001	31132	+
------	-------	-------	---

Inter	31211	31309	+
-------	-------	-------	---

Inter	31406	31522	+
-------	-------	-------	---

Intr	31618	31713	+
------	-------	-------	---

Intr	31848	31960	+
------	-------	-------	---

>3176693 /28669

```
len = 5624 nex = 10
```

Term	36200	35812	-
------	-------	-------	---

Intr	36523	36290	-
------	-------	-------	---

Intr	37268	36993	-
------	-------	-------	---

Intr	37641	37365	-
------	-------	-------	---

55	Intr	38211	37913	-
----	------	-------	-------	---

Intr	38769	38584	-
------	-------	-------	---

Intr	39653	39444	-
------	-------	-------	---

Intr	39839	39776	-
------	-------	-------	---

Intr	40136	39990	-
------	-------	-------	---

60	Init	40542	40305	-
----	------	-------	-------	---

1120

	>3176693	/12338			
5	len =	1346	nex =	2	
	Term	8005	7949	-	0
	Init	8926	8665	-	0
10	>3176694	/10539			
	len =	704	nex =	1	
	Sngl	11287	11990	+	0
15	>3176694	/206065			
	len =	1371	nex =	3	
20	Term	19165	18447	-	0
	Intr	19610	19409	-	0
	Init	19817	19698	-	0
	>3176694	/3965			
25	len =	356	nex =	1	
	Sngl	29025	28693	-	0
30	>3176695	/39581			
	len =	3411	nex =	7	
	Term	8155	7742	-	0
	Intr	8402	8251	-	0
35	Intr	8648	8490	-	0
	Intr	9014	8762	-	0
	Intr	9217	9101	-	0
	Intr	9356	9294	-	0
	Init	11152	10747	-	0
40	>3176695	/25260			
	len =	897	nex =	3	
45	Init	26015	26131	+	0
	Intr	26251	26367	+	0
	Term	26507	26911	+	0
50	>3176695	/34123			
	len =	215	nex =	1	
	Sngl	38577	38363	-	0
55	>3176695	/28545			
	len =	889	nex =	2	
60	Term	44351	43984	-	0
	Init	44872	44663	-	0

1121

>3176695 /32366

5	len =	1810	nex =	5	
	Init	45733	45892	+	0
	Intr	46308	46477	+	0
	Intr	46559	46750	+	0
	Intr	46829	47178	+	0
10	Term	47280	47536	+	0

>3176695 /18004

15	len =	119	nex =	1	
	Sngl	75443	75561	+	0

>3176701 /37338

20	len =	2170	nex =	6	
	Init	23063	23245	+	0
	Intr	23800	24134	+	0
	Intr	24237	24313	+	0
25	Intr	24402	24462	+	0
	Intr	24573	24717	+	0
	Term	24845	25227	+	0

>3176701 /39441

30	len =	1243	nex =	3	
	Init	84984	85415	+	0
	Intr	85710	85873	+	0
35	Term	86105	86226	+	0

>3176701 /38836

40	len =	490	nex =	2	
	Init	85708	85873	+	0
	Term	86105	86195	+	0

>3184270 /26825

45	len =	1709	nex =	2	
	Term	46892	46301	-	0
	Init	48009	47436	-	0

>3184270 /13775

	len =	2935	nex =	9	
55	Init	52068	52252	+	0
	Intr	52654	52722	+	0
	Intr	53313	53438	+	0
	Intr	53646	53728	+	0
	Intr	53832	53907	+	0
60	Intr	53982	54158	+	0

1122

Intr	54241	54369	+	0
Intr	54444	54582	+	0
Term	54778	55002	+	0

5 >3184270 /12948

len = 399 nex = 1

10 Sngl 77713 77315 - 0

>3184270 /13304

len = 2294 nex = 11

15 Init 80046 80143 + 0

Intr 80300 80371 + 0

Intr 80466 80521 + 0

Intr 80725 80804 + 0

Intr 80900 80982 + 0

20 Intr 81099 81176 + 0

Intr 81270 81338 + 0

Intr 81422 81490 + 0

Intr 81567 81613 + 0

25 Intr 81699 81789 + 0

Term 81859 82119 + 0

>3184270 /108825

len = 1969 nex = 10

30 Init 80300 80371 + 0

Intr 80466 80521 + 0

Intr 80725 80804 + 0

Intr 80900 80982 + 0

35 Intr 81099 81176 + 0

Intr 81270 81338 + 0

Intr 81422 81490 + 0

Intr 81567 81613 + 0

40 Intr 81699 81789 + 0

Term 81859 82035 + 0

>3193282 /13633

len = 1915 nex = 4

45 Term 16811 16369 - 0

Intr 17513 17355 - 0

Intr 17971 17940 - 0

Init 18283 18065 - 0

50 >3193282 /20206

len = 2050 nex = 6

55 Init 35088 35203 + 0

Intr 35751 35937 + 0

Intr 36023 36179 + 0

Intr 36287 36414 + 0

60 Intr 36457 36643 + 0

Term 36865 37128 + 0

1123

>3193282 /145583

5	len =	614	nex =	2	
	Init	36528	36643	+	0
	Term	36865	37141	+	0

>3193282 /19681

10	len =	2692	nex =	10	
	Term	39121	39088	-	0
	Intr	40307	40078	-	0
15	Intr	40438	40381	-	0
	Intr	40596	40547	-	0
	Intr	40783	40726	-	0
	Intr	40918	40872	-	0
	Intr	41132	40997	-	0
20	Intr	41298	41224	-	0
	Intr	41506	41441	-	0
	Init	41770	41692	-	0

>3193282 /25565

25	len =	1738	nex =	7	
	Init	51343	51536	+	0
	Intr	51738	51895	+	0
30	Intr	51986	52106	+	0
	Intr	52242	52366	+	0
	Intr	52446	52559	+	0
	Intr	52655	52822	+	0
35	Term	52894	53080	+	0

>3193282 /23194

	len =	717	nex =	2	
40	Init	63968	64230	+	0
	Term	64331	64684	+	0

>3193282 /17524

45	len =	707	nex =	1	
	Sngl	68554	67848	-	0

>3193282 /19543

50	len =	869	nex =	1	
	Sngl	68626	67758	-	0

>3193282 /15281

55	len =	1873	nex =	3	
	Term	73831	73399	-	0
60	Intr	74170	73932	-	0

2025 RELEASE UNDER E.O. 14176

1124

	Init	75271	74778	-	0
	>3193305 /40560				
5	len =	2320	nex =	6	
	Term	10008	9622	-	0
	Intr	10234	10083	-	0
10	Intr	10745	10346	-	0
	Intr	10952	10836	-	0
	Intr	11528	11466	-	0
	Init	11941	11625	-	0
	>3193305 /29200				
15	len =	986	nex =	1	
	Sngl	30156	29171	-	0
20	>3193305 /18253				
	len =	670	nex =	1	
	Sngl	31064	31727	+	0
25	>3193311 /19785				
	len =	2718	nex =	10	
30	Init	27978	28299	+	0
	Intr	28407	28512	+	0
	Intr	28610	28670	+	0
	Intr	28784	29064	+	0
	Intr	29273	29412	+	0
35	Intr	29489	29563	+	0
	Intr	29673	29763	+	0
	Intr	29932	30135	+	0
	Intr	30249	30293	+	0
	Term	30371	30695	+	0
40	>3193311 /4058				
	len =	1211	nex =	4	
45	Init	30962	31056	+	0
	Intr	31136	31208	+	0
	Intr	31291	31429	+	0
	Term	31513	31699	+	0
50	>3193311 /12983				
	len =	2068	nex =	4	
	Init	57272	57337	+	0
55	Intr	57712	58588	+	0
	Intr	58663	58797	+	0
	Term	58889	59339	+	0
60	>3193311 /32470				

1125

	len =	610	nex =	1	
	Sngl	63745	63139	-	0
5	>3193311	/33482			
	len =	1692	nex =	7	
10	Term	64707	64353	-	0
	Intr	64872	64795	-	0
	Intr	65002	64961	-	0
	Intr	65233	65072	-	0
	Intr	65434	65319	-	0
15	Intr	65755	65682	-	0
	Init	66044	65896	-	0
	>3193311	/102945			
20	len =	2350	nex =	5	
	Term	66929	66852	-	0
	Intr	67385	67310	-	0
	Intr	67569	67521	-	0
25	Intr	68289	68130	-	0
	Init	68762	68559	-	0
	>3193311	/13263			
30	len =	2770	nex =	6	
	Term	66929	66852	-	0
	Intr	67385	67310	-	0
	Intr	67569	67521	-	0
35	Intr	68289	68130	-	0
	Intr	68729	68559	-	0
	Init	69246	68838	-	0
	>3193311	/21877			
40	len =	1450	nex =	2	
	Init	8494	8695	+	0
	Term	8783	9939	+	0
45	>3201608	/30206			
	len =	916	nex =	3	
50	Init	13538	13733	+	0
	Intr	13827	13875	+	0
	Term	14061	14453	+	0
	>3201608	/36378			
55	len =	337	nex =	2	
	Init	13539	13733	+	0
	Term	13827	13875	+	0
60	>3201608	/38967			

1126

	len =	1123	nex =	3	
5	Init	15910	16091	+	0
	Intr	16178	16226	+	0
	Term	16396	17032	+	0
	>3201608 /142926				
10	len =	446	nex =	1	
	Sngl	15916	16359	+	0
	>3201608 /33579				
15	len =	430	nex =	1	
	Sngl	16612	17032	+	0
20	>3201608 /7866				
	len =	2272	nex =	4	
25	Init	22124	22360	+	0
	Intr	22689	22844	+	0
	Intr	23414	23562	+	0
	Term	23794	24395	+	0
	>3201608 /125631				
30	len =	956	nex =	3	
	Init	36898	37143	+	0
	Intr	37222	37407	+	0
35	Term	37534	37853	+	0
	>3201608 /118068				
40	len =	680	nex =	1	
	Sngl	52418	52003	-	0
	>3201608 /34360				
45	len =	2916	nex =	10	
	Init	53417	53661	+	0
	Intr	53871	53956	+	0
	Intr	54282	54488	+	0
50	Intr	54612	54717	+	0
	Intr	54802	54906	+	0
	Intr	54980	55176	+	0
	Intr	55376	55535	+	0
	Intr	55661	55756	+	0
55	Intr	55848	55942	+	0
	Term	56072	56332	+	0
	>3201608 /18932				
60	len =	3685	nex =	12	

1127

	Term	56943	56652	-	0
	Intr	57107	57006	-	0
5	Intr	57319	57236	-	0
	Intr	57540	57445	-	0
	Intr	57666	57616	-	0
	Intr	57901	57761	-	0
	Intr	58146	58090	-	0
10	Intr	58297	58223	-	0
	Intr	58658	58386	-	0
	Intr	58800	58741	-	0
	Intr	59106	59014	-	0
	Init	60336	60218	-	0

```
15  >3201608      /36479
      len =      3770    nex =      11
```

20	Term	57107	57006	-	0
	Intr	57319	57236	-	0
	Intr	57540	57445	-	0
	Intr	57666	57616	-	0
	Intr	57901	57761	-	0
25	Intr	58146	58090	-	0
	Intr	58297	58223	-	0
	Intr	58658	58386	-	0
	Intr	58800	58741	-	0
	Intr	59106	59014	-	0
	Init	60381	60218	-	0

```

30      >3201608      /35095
      len =      2050      nex =      9

```

35	Term	83717	83480	-	0
	Intr	83885	83817	-	0
	Intr	84083	83991	-	0
	Intr	84260	84169	-	0
40	Intr	84508	84428	-	0
	Intr	84761	84666	-	0
	Intr	84937	84841	-	0
	Intr	85083	85034	-	0
	Init.	85526	85162	-	0

```
45  >3212102      /36315
      len =      1724    nex =      8
```

50	Init	20146	20243	+	0
	Intr	20513	20640	+	0
	Intr	20801	20899	+	0
	Intr	20991	21149	+	0
	Intr	21233	21319	+	0
55	Intr	21394	21485	+	0
	Intr	21563	21637	+	0
	Term	21735	21869	+	0

>3212102 /124895

```
60      len =      930      nex =      2
```

1128

	Term	23160	22726	-	0
	Init	23655	23565	-	0
5	>3212102	/7088			
	len =	1376	nex =	2	
	Term	23160	22971	-	0
10	Init	24346	23565	-	0
	>3212102	/42528			
	len =	1189	nex =	2	
15	Init	54373	54915	+	0
	Term	55009	55561	+	0
	>3212102	/34748			
20	len =	631	nex =	1	
	Sngl	71980	71350	-	0
25	>3212102	/25849			
	len =	437	nex =	1	
	Sngl	71980	71544	-	0
30	>3212846	/30161			
	len =	1396	nex =	6	
35	Term	100365	100258	-	0
	Intr	100551	100455	-	0
	Intr	100842	100748	-	0
	Intr	101023	100927	-	0
	Intr	101352	101107	-	0
40	Init	101653	101420	-	0
	>3212846	/24370			
	len =	2551	nex =	8	
45	Term	14963	14742	-	0
	Intr	15363	15285	-	0
	Intr	15522	15476	-	0
	Intr	15716	15617	-	0
50	Intr	15974	15874	-	0
	Intr	16145	16059	-	0
	Intr	16807	16677	-	0
	Init	17292	17119	-	0
55	>3212846	/36599			
	len =	598	nex =	2	
	Term	16807	16695	-	0
60	Init	17292	17119	-	0

2025-06-10 10:00:00

>3212846 /106170

5	len =	1639	nex =	6	
	Term	19624	19275	-	0
	Intr	19851	19773	-	0
	Intr	20034	19988	-	0
10	Intr	20229	20130	-	0
	Intr	20408	20308	-	0
	Init	20607	20521	-	0

>3212846 /38206

15	len =	2304	nex =	6	
	Term	28397	28158	-	0
	Intr	28537	28479	-	0
	Intr	28784	28635	-	0
20	Intr	28981	28865	-	0
	Intr	29343	29149	-	0
	Init	30461	30043	-	0

>3212846 /482

25	len =	1849	nex =	7	
	Term	36128	35534	-	0
	Intr	36364	36203	-	0
30	Intr	36609	36541	-	0
	Intr	36747	36700	-	0
	Intr	36942	36840	-	0
	Intr	37120	37041	-	0
35	Init	37271	37200	-	0

>3212846 /36111

	len =	1817	nex =	8	
40	Term	52320	52252	-	0
	Intr	52529	52442	-	0
	Intr	52740	52631	-	0
	Intr	52934	52833	-	0
	Intr	53111	53037	-	0
45	Intr	53340	53207	-	0
	Intr	53778	53541	-	0
	Init	54068	53940	-	0

>3212846 /10293

50	len =	2263	nex =	8	
	Term	52320	52036	-	0
	Intr	52529	52442	-	0
55	Intr	52740	52631	-	0
	Intr	52934	52833	-	0
	Intr	53111	53037	-	0
	Intr	53340	53207	-	0
	Intr	53778	53541	-	0
60	Init	54153	53940	-	0

1130

>3212846 /26137

5	len =	1882	nex =	8	
	Init	54528	54721	+	0
	Intr	54809	54924	+	0
	Intr	55061	55156	+	0
	Intr	55233	55328	+	0
10	Intr	55472	55550	+	0
	Intr	55791	55837	+	0
	Intr	55936	55995	+	0
	Term	56146	56409	+	0

15 >3212846 /23761

	len =	1826	nex =	1	
20	Sngr	58250	58411	+	0

>3212846 /2450

	len =	1830	nex =	2	
25	Init	57184	58160	+	0
	Term	58250	58415	+	0

>3212846 /24003

30	len =	2050	nex =	3	
	Init	58759	58825	+	0
	Intr	59065	59223	+	0
	Term	60088	60794	+	0

35 >3212846 /23722

	len =	2058	nex =	3	
40	Init	58759	58825	+	0
	Intr	59065	59223	+	0
	Term	60088	60804	+	0

>3212846 /29536

45	len =	3420	nex =	13	
	Init	71959	72157	+	0
	Intr	72507	72569	+	0
50	Intr	72658	72741	+	0
	Intr	72831	72938	+	0
	Intr	73127	73299	+	0
	Intr	73393	73504	+	0
	Intr	73614	73694	+	0
55	Intr	73795	73857	+	0
	Intr	73975	74076	+	0
	Intr	74249	74373	+	0
	Intr	74619	74829	+	0
	Intr	74912	74986	+	0
60	Term	75078	75378	+	0

000000-000000

1131

>3212846 /4611

	len =	2878	nex =	11	
5	Init	72071	72157	+	0
	Intr	72507	72569	+	0
	Intr	72658	72741	+	0
	Intr	72831	72938	+	0
10	Intr	73127	73299	+	0
	Intr	73393	73504	+	0
	Intr	73614	73694	+	0
	Intr	73795	73857	+	0
	Intr	73975	74076	+	0
15	Intr	74249	74373	+	0
	Term	74619	74829	+	0

>3212846 /36813

20	len =	1667	nex =	5	
	Term	79812	79309	-	0
	Intr	79978	79899	-	0
	Intr	80211	80154	-	0
25	Intr	80392	80295	-	0
	Init	80975	80496	-	0

>3212846 /17409

30	len =	1733	nex =	5	
	Term	79812	79361	-	0
	Intr	79978	79899	-	0
	Intr	80211	80154	-	0
35	Intr	80392	80295	-	0
	Init	81093	80496	-	0

>3212846 /30978

40	len =	1706	nex =	4	
	Term	79812	79388	-	0
	Intr	80211	79899	-	0
	Intr	80392	80295	-	0
45	Init	81093	80496	-	0

>3212846 /6950

	len =	1390	nex =	1	
50	Sngl	83877	85266	+	0

>3212846 /17908

55	len =	2271	nex =	8	
	Init	93341	93451	+	0
	Intr	93759	93932	+	0
	Intr	94012	94112	+	0
60	Intr	94210	94321	+	0

1132

	Intr	94487	94567	+	0
	Intr	94665	94842	+	0
	Intr	94936	95270	+	0
	Term	95369	95611	+	0
5	>3228389 /117479				
	len =	1270	nex =	2	
10	Init	25417	26084	+	0
	Term	26445	26682	+	0
	>3228389 /7227				
15	len =	1347	nex =	2	
	Init	25417	26084	+	0
	Term	26445	26763	+	0
20	>3228389 /15453				
	len =	1292	nex =	2	
	Term	42250	42036	-	0
25	Init	42669	42513	-	0
	>3228389 /2617				
	len =	1296	nex =	2	
30	Term	42250	42033	-	0
	Init	42669	42513	-	0
	>3228389 /42666				
35	len =	1771	nex =	4	
	Init	45914	46685	+	0
	Intr	46846	46967	+	0
40	Intr	47100	47351	+	0
	Term	47455	47684	+	0
	>3228389 /39286				
45	len =	1930	nex =	5	
	Term	55070	54825	-	0
	Intr	55329	55246	-	0
	Intr	55816	55703	-	0
50	Intr	56104	56001	-	0
	Init	56745	56456	-	0
	>3228389 /18771				
55	len =	320	nex =	1	
	Sngl	57320	57001	-	0
60	>3228389 /749				

000000-000000

1133

len = 1712 nex = 7
 Init 59020 59117 + 0
 Intr 59214 59250 + 0
 5 Intr 59369 59409 + 0
 Intr 59889 60119 + 0
 Intr 60204 60306 + 0
 Intr 60397 60464 + 0
 Term 60563 60731 + 0
 10 >3228389 /34212
 len = 1113 nex = 2
 15 Term 71736 71610 - 0
 Init 72722 72553 - 0
 >3228389 /22723
 20 len = 1056 nex = 1
 Sngl 72763 72553 - 0
 >3228389 /3279
 25 len = 1656 nex = 6
 Init 76223 76387 + 0
 Intr 76478 76588 + 0
 30 Intr 76756 76848 + 0
 Intr 77250 77399 + 0
 Intr 77484 77542 + 0
 Term 77636 77878 + 0
 35 >3228389 /35408
 len = 1760 nex = 2
 Init 88411 88892 + 0
 40 Term 89156 90170 + 0
 >3236234 /42701
 len = 2562 nex = 5
 45 Init 14714 14911 + 0
 Intr 15858 15964 + 0
 Intr 16057 16189 + 0
 Intr 16311 16602 + 0
 50 Term 16755 17275 + 0
 >3236234 /116968
 len = 826 nex = 1
 55 Sngl 20794 21619 + 0
 >3236234 /18641
 60 len = 814 nex = 1

009100.9255550

1134

	Sngl	21539	20726	-	0
	>3236234	/34743			
5	len =	1033	nex =	3	
	Term	30515	30312	-	0
	Intr	30858	30659	-	0
10	Init	31344	31168	-	0
	>3236234	/23114			
	len =	1095	nex =	3	
15	Term	30515	30314	-	0
	Intr	30858	30659	-	0
	Init	31275	31168	-	0
20	>3236234	/19581			
	len =	2170	nex =	4	
	Init	35817	35886	+	0
25	Intr	36023	36416	+	0
	Intr	36938	37130	+	0
	Term	37242	37510	+	0
	>3236234	/38101			
	len =	1484	nex =	4	
	Init	35822	35886	+	0
	Intr	36023	36416	+	0
35	Intr	36478	37130	+	0
	Term	37242	37305	+	0
	>3236234	/25785			
	len =	1487	nex =	4	
	Init	35822	35886	+	0
	Intr	36023	36416	+	0
	Intr	36938	37130	+	0
45	Term	37242	37308	+	0
	>3236234	/13797			
	len =	1695	nex =	7	
50	Term	51666	51431	-	0
	Intr	51905	51782	-	0
	Intr	52049	51988	-	0
	Intr	52228	52131	-	0
55	Intr	52342	52306	-	0
	Intr	52698	52427	-	0
	Init	53125	52779	-	0
	>3236234	/28529			
60					

1135

len = 529 nex = 1

Sngl 87056 87584 + 0

5 >3236479 /10950

len = 1705 nex = 2

Init 28178 28532 + 0

10 Term 29443 29882 + 0

>3236479 /30441

len = 285 nex = 1

15 Sngl 28227 28511 + 0

>3236479 /143197

20 len = 439 nex = 1

Sngl 38868 39306 + 0

>3236479 /32070

25 len = 2470 nex = 4

Term 89296 89181 - 0

Intr 89764 89651 - 0

30 Intr 89981 89841 - 0

Init 91120 90101 - 0

>3241916 /143222

35 len = 1330 nex = 6

Term 17392 17118 - 0

Intr 17601 17497 - 0

Intr 17779 17689 - 0

40 Intr 17948 17886 - 0

Intr 18263 18214 - 0

Init 18446 18359 - 0

>3241916 /7891

45 len = 1510 nex = 2

Init 19560 19940 + 0

Term 20202 21066 + 0

50 >3241916 /105027

len = 167 nex = 1

55 Sngl 42357 42523 + 0

>3241916 /2036

len = 1660 nex = 9

60

000000-000000

1136

	Term	42888	42671	-	0
	Intr	43123	42976	-	0
	Intr	43297	43247	-	0
	Intr	43511	43397	-	0
5	Intr	43721	43603	-	0
	Intr	43846	43802	-	0
	Intr	43979	43940	-	0
	Intr	44199	44093	-	0
	Init	44330	44290	-	0
10	>3241916 /12974				
	len =	1882	nex =	5	
15	Term	52467	52276	-	0
	Intr	52920	52744	-	0
	Intr	53121	53006	-	0
	Intr	53292	53235	-	0
	Init	53637	53395	-	0
20	>3241917 /148899				
	len =	1424	nex =	2	
25	Term	12645	12184	-	0
	Init	13607	13335	-	0
	>3241917 /3896				
30	len =	1486	nex =	2	
	Term	12645	12218	-	0
	Init	13703	13335	-	0
35	>3241917 /152864				
	len =	319	nex =	1	
40	Sngl	13703	13385	-	0
	>3241917 /119748				
	len =	794	nex =	2	
45	Term	28884	28539	-	0
	Init	29332	28954	-	0
	>3241917 /23920				
50	len =	1487	nex =	4	
	Init	37712	38242	+	0
	Intr	38331	38575	+	0
	Intr	38659	38777	+	0
55	Term	38871	39198	+	0
	>3241917 /148141				
60	len =	1491	nex =	5	

2025 RELEASE UNDER E.O. 14176

1137

	Init	44373	44517	+	0
	Intr	44592	44617	+	0
	Intr	44716	44805	+	0
	Intr	44885	44973	+	0
5	Term	45057	45296	+	0

>3241917 /37636

10	len =	939	nex =	5	
	Init	44355	44517	+	0
	Intr	44592	44617	+	0
	Intr	44716	44805	+	0
	Intr	44885	44973	+	0
15	Term	45057	45293	+	0

>3241917 /34484

20	len =	490	nex =	1	
	Sngl	51169	51654	+	0

>3241917 /30313

25	len =	1235	nex =	4	
	Init	57202	57315	+	0
	Intr	57401	57469	+	0
	Intr	57768	57870	+	0
30	Term	58209	58436	+	0

>3241920 /19080

35	len =	1218	nex =	1	
	Sngl	45667	46884	+	0

>3241921 /42897

40	len =	475	nex =	1	
	Sngl	26675	26201	-	0

>3241921 /33053

45	len =	555	nex =	1	
	Sngl	35941	36495	+	0

50	>3241922	/29124			
	len =	1276	nex =	4	

	Init	18356	18395	+	0
	Intr	18580	18665	+	0
	Intr	18750	18873	+	0
	Term	19391	19631	+	0

60	>3241922	/34761			
----	----------	--------	--	--	--

2025 RELEASE UNDER E.O. 14176

1138

	len =	420	nex =	1	
	Sngl	288	707	+	0
5	>3241922	/13557			
	len =	1016	nex =	4	
10	Init	47961	48011	+	0
	Intr	48229	48408	+	0
	Intr	48587	48661	+	0
	Term	48747	48972	+	0
	>3241922	/29542			
15	len =	1845	nex =	4	
	Term	49432	49038	-	0
	Intr	49649	49523	-	0
20	Intr	50374	50138	-	0
	Init	50882	50532	-	0
	>3241922	/100927			
25	len =	391	nex =	1	
	Sngl	58311	58701	+	0
	>3241922	/34244			
30	len =	166	nex =	1	
	Sngl	84414	84579	+	0
35	>3241923	/40576			
	len =	765	nex =	1	
40	Sngl	16439	17203	+	0
	>3241923	/8215			
	len =	1690	nex =	8	
45	Term	21000	20843	-	0
	Intr	21226	21110	-	0
	Intr	21370	21308	-	0
	Intr	21576	21479	-	0
	Intr	21730	21665	-	0
50	Intr	22067	21967	-	0
	Intr	22235	22163	-	0
	Init	22526	22444	-	0
	>3241923	/11699			
55	len =	600	nex =	1	
	Sngl	24791	24991	+	0
60	>3241923	/7822			

1139

	len =	595	nex =	3	
	Term	26064	25987	-	0
5	Intr	26320	26160	-	0
	Init	26559	26419	-	0
	>3241923	/36757			
10	len =	1390	nex =	2	
	Term	26064	25685	-	0
	Init	26320	26160	-	0
15	>3241923	/20036			
	len =	734	nex =	2	
	Term	29311	28952	-	0
20	Init	29685	29399	-	0
	>3241923	/20973			
	len =	1537	nex =	3	
25	Init	33738	33851	+	0
	Intr	34586	34851	+	0
	Term	34945	35274	+	0
30	>3241923	/11691			
	len =	1472	nex =	2	
	Init	36556	36821	+	0
35	Term	36913	37365	+	0
	>3241923	/9208			
	len =	510	nex =	2	
40	Init	36555	36821	+	0
	Term	36913	37064	+	0
	>3241923	/20822			
45	len =	715	nex =	1	
	Sngl	40595	41309	+	0
50	>3241923	/114182			
	len =	399	nex =	1	
	Sngl	40736	41127	+	0
55	>3241923	/27649			
	len =	1630	nex =	5	
60	Init	53569	53872	+	0

1140

Intr	54069	54212	+	0
Intr	54361	54429	+	0
Intr	54523	54701	+	0
Term	54800	55196	+	0

5

>3241923 /26907

len = 1617 nex = 5

10	Init	53602	53872	+	0
	Intr	54069	54212	+	0
	Intr	54361	54429	+	0
	Intr	54523	54701	+	0
	Term	54800	55218	+	0

15

>3241923 /31173

len = 440 nex = 1

20	Sngl	60193	59754	-	0
----	------	-------	-------	---	---

>3241923 /4349

len = 1995 nex = 6

25	Term	8413	7946	-	0
	Intr	8649	8527	-	0
	Intr	8936	8739	-	0
	Intr	9218	9120	-	0
30	Intr	9456	9304	-	0
	Init	9940	9737	-	0

>3241924 /4012

35	len =	407	nex =	0
----	-------	-----	-------	---

>3241924 /2034

len = 574 nex = 1

40	Sngl	2980	3553	+	0
----	------	------	------	---	---

>3241924 /41458

45	len =	2372	nex =	10
----	-------	------	-------	----

	Term	32689	32396	-	0
	Intr	32948	32773	-	0
	Intr	33091	33036	-	0
50	Intr	33270	33226	-	0
	Intr	33426	33360	-	0
	Intr	33636	33525	-	0
	Intr	33778	33734	-	0
	Intr	33948	33876	-	0
55	Intr	34113	34039	-	0
	Init	34286	34212	-	0

>3241924 /38213

60	len =	1734	nex =	5
----	-------	------	-------	---

1141

	Term	36042	35649	-	0
	Intr	36432	36137	-	0
	Intr	36674	36516	-	0
5	Intr	37042	36742	-	0
	Init	37382	37126	-	0
	>3241924 /31303				
10	len =	1072	nex =	3	
	Init	45856	45947	+	0
	Intr	46181	46247	+	0
	Term	46336	46715	+	0
15	>3241924 /14377				
	len =	768	nex =	3	
20	Term	51633	51312	-	0
	Intr	51804	51711	-	0
	Init	52079	51889	-	0
25	>3241924 /38528				
	len =	1902	nex =	6	
	Term	51633	51293	-	0
	Intr	51804	51711	-	0
30	Intr	52142	51889	-	0
	Intr	52355	52232	-	0
	Intr	52541	52448	-	0
	Init	53194	53036	-	0
35	>3241924 /7358				
	len =	490	nex =	1	
	Sngr	58044	58525	+	0
40	>3241924 /106913				
	len =	1215	nex =	2	
45	Init	77459	77620	+	0
	Term	78380	78673	+	0
	>3241925 /10394				
50	len =	2076	nex =	6	
	Term	49380	48998	-	0
	Intr	49603	49472	-	0
	Intr	49732	49678	-	0
55	Intr	50653	50482	-	0
	Intr	50890	50743	-	0
	Init	51073	50979	-	0
	>3241926 /25388				
60					

1142

	len =	944	nex =	2	
	Init	12191	12401	+	0
	Term	12491	12805	+	0
5	>3241926	/41509			
	len =	1870	nex =	4	
	Init	13521	13588	+	0
	Intr	13687	14214	+	0
	Intr	14299	14486	+	0
	Term	14580	14889	+	0
15	>3241926	/13875			
	len =	1717	nex =	5	
	Init	13173	13425	+	0
	Intr	13521	13588	+	0
	Intr	13687	14214	+	0
	Intr	14299	14486	+	0
	Term	14580	14889	+	0
25	>3241926	/18612			
	len =	2833	nex =	7	
	Term	15300	14853	-	0
	Intr	15460	15363	-	0
	Intr	15616	15557	-	0
	Intr	15814	15695	-	0
	Intr	16062	15916	-	0
	Intr	16256	16155	-	0
35	Init	17421	16839	-	0
	>3241926	/8268			
	len =	1345	nex =	6	
	Init	6075	6204	+	0
	Intr	6527	6560	+	0
	Intr	6653	6825	+	0
	Intr	6914	6965	+	0
	Intr	7061	7128	+	0
	Term	7215	7419	+	0
	>3241926	/206563			
	len =	1570	nex =	5	
	Term	74328	73909	-	0
	Intr	74617	74464	-	0
	Intr	74837	74697	-	0
	Intr	74994	74923	-	0
	Init	75192	75080	-	0
	>3241927	/4309			
	len =	757	nex =	0	

1143

5 >3241927 /32995
 len = 2052 nex = 0
 10 >3241939 /27423
 len = 1306 nex = 3
 10 Init 2469 2639 + 0
 Intr 2957 3312 + 0
 Term 3400 3774 + 0
 15 >3241939 /31388
 len = 970 nex = 2
 Init 2957 3312 + 0
 Term 3400 3649 + 0
 20 >3241939 /41130
 len = 1254 nex = 3
 25 Term 28798 28404 - 0
 Intr 29363 29256 - 0
 Init 29657 29438 - 0
 30 >3242700 /120446
 len = 358 nex = 1
 Sngl 37693 38050 + 0
 35 >3242700 /25463
 len = 1390 nex = 2
 Term 93971 93636 - 0
 40 Init 95019 94055 - 0
 >3242700 /26006
 len = 1391 nex = 3
 45 Term 93971 93636 - 0
 Intr 94189 94055 - 0
 Init 95026 94632 - 0
 50 >3242970 /34126
 len = 1718 nex = 5
 Init 61864 62009 + 0
 55 Intr 62392 62498 + 0
 Intr 62588 62788 + 0
 Init 63185 63249 + 0
 Term 63334 63581 + 0
 60 >3242970 /99825

2025 RELEASE UNDER E.O. 14176

1144

	len =	1815	nex =	5	
	Init	61865	62009	+	0
5	Intr	62392	62498	+	0
	Intr	62588	62788	+	0
	Intr	63185	63249	+	0
	Term	63334	63679	+	0
10	>3243214	/24559			
	len =	1330	nex =	3	
	Term	22739	22559	-	0
15	Intr	23241	23208	-	0
	Init	23516	23386	-	0
	>3249094	/9228			
20	len =	951	nex =	1	
	Sngl	12849	11899	-	0
	>3249094	/37124			
25	len =	1184	nex =	1	
	Sngl	17368	16185	-	0
30	>3249094	/27210			
	len =	3696	nex =	13	
	Init	25520	25687	+	0
35	Intr	26293	26400	+	0
	Intr	26495	26687	+	0
	Intr	26785	27043	+	0
	Intr	27217	27321	+	0
	Intr	27408	27494	+	0
40	Intr	27577	27634	+	0
	Intr	28142	28209	+	0
	Intr	28295	28390	+	0
	Intr	28484	28531	+	0
	Intr	28693	28722	+	0
45	Intr	28807	28923	+	0
	Term	29021	29215	+	0
	>3249094	/10388			
50	len =	587	nex =	1	
	Sngl	30050	30636	+	0
	>3249094	/1854			
55	len =	792	nex =	3	
	Term	36628	36312	-	0
	Intr	36865	36734	-	0
60	Init	37103	36974	-	0

1145

>3249094 /31022

5	len =	1314	nex =	3	
	Init	51260	51655	+	0
	Intr	51724	51842	+	0
	Term	51917	52573	+	0

10 >3249094 /110171

	len =	2770	nex =	8	
15	Term	53358	53234	-	0
	Intr	53539	53476	-	0
	Intr	53856	53789	-	0
	Intr	54097	53939	-	0
	Intr	54429	54377	-	0
	Intr	54692	54595	-	0
20	Intr	54861	54794	-	0
	Init	55039	54958	-	0

>3249094 /7559

25	len =	2785	nex =	11	
	Init	56298	56517	+	0
	Intr	56641	56706	+	0
	Intr	56793	56867	+	0
30	Intr	56947	57049	+	0
	Intr	57135	57181	+	0
	Intr	57289	57346	+	0
	Intr	57470	57519	+	0
	Intr	57663	57719	+	0
35	Intr	57793	57905	+	0
	Intr	58003	58080	+	0
	Term	59023	59082	+	0

>3250673 /3164

40	len =	734	nex =	1	
	Sngl	17027	17760	+	0

45 >3250673 /17437

	len =	1613	nex =	3	
50	Init	83877	84182	+	0
	Intr	84570	84691	+	0
	Term	84930	85489	+	0

>3250673 /34882

55	len =	1175	nex =	3	
	Term	94732	94604	-	0
	Intr	95043	94819	-	0
	Init	95309	95132	-	0

60

2025 RELEASE UNDER E.O. 14176

1146

	>3252804	/40603			
	len =	4015	nex =	10	
5	Term	17729	17324	-	0
	Intr	18030	17844	-	0
	Intr	18354	18243	-	0
	Intr	18716	18462	-	0
	Intr	19017	18946	-	0
10	Intr	19896	19802	-	0
	Intr	20112	19995	-	0
	Intr	20703	20628	-	0
	Intr	21077	20991	-	0
15	Init	21338	21278	-	0
	>3252804	/14216			
	len =	1608	nex =	7	
20	Term	72843	72609	-	0
	Intr	73012	72966	-	0
	Intr	73182	73098	-	0
	Intr	73389	73271	-	0
	Intr	73624	73475	-	0
25	Intr	73849	73709	-	0
	Init	74216	73929	-	0
	>3269280	/15604			
30	len =	971	nex =	2	
	Term	13730	13454	-	0
	Init	14424	14146	-	0
35	>3269280	/927			
	len =	1073	nex =	2	
	Term	13730	13356	-	0
40	Init	14428	14146	-	0
	>3269280	/2944			
45	len =	1694	nex =	2	
	Init	15972	16829	+	0
	Term	17015	17665	+	0
	>3269280	/13072			
50	len =	670	nex =	1	
	Sngl	40249	39588	-	0
55	>3269280	/17584			
	len =	594	nex =	1	
60	Sngl	40252	39659	-	0

20250927 14:22:55

1147

>3269280 /10313

```
len = 2310    nex = 2
```

5	Term	54740	54478	-	0
	Init	56787	55952	-	0

>3269280 /35749

```
10      len = 1969      nex = 2
```

Term	67347	66835	-	0
Init	68803	68154	-	0

15 >3269280 /39662

```
len = 1890    nex = 5
```

	Init	77267	77439	+	0
20	Intr	77520	77574	+	0
	Intr	78008	78078	+	0
	Intr	78194	78302	+	0
	Term	78818	79156	+	0

25 >3269280 /40387

```
len =      893    nex =      2
```

	Init	81954	82059	+	0
30	Term	82190	82846	+	0

>3281847 /18361

```
len = 476    nex = 2
```

Term	10602	10258	-	0
Init	10733	10682	-	0

>3281847 /25793

```

40      len =      2313      nex =      6

```

Init	31587	31715	+	0
Intr	31812	31876	+	0

45	Intr	32278	32340	+	0
	Intr	32526	32577	+	0
	Intr	32672	32722	+	0
	Term	33024	33293	+	0

50 >3281847 /34408

```
len = 1487    nex = 5
```

55	Init	39703	39769	+	0
	Intr	40039	40096	+	0
	Intr	40195	40303	+	0
	Intr	40413	40685	+	0
	Term	40780	41189	+	0

60 >3281847 /33868

1149

	Term	59522	59409	-	0
	Intr	59658	59614	-	0
	Intr	59899	59756	-	0
5	Intr	60368	60000	-	0
	Intr	61028	60447	-	0
	Intr	61250	61122	-	0
	Init	61853	61594	-	0
10	>3282170	/30230			
	len =	870	nex =	1	
	Sngl	70042	70911	+	0
15	>3282170	/32125			
	len =	1584	nex =	1	
20	Sngl	84969	86552	+	0
	>3282170	/2939			
	len =	310	nex =	1	
25	Sngl	86228	86537	+	0
	>3282170	/25426			
30	len =	458	nex =	1	
	Sngl	88205	88662	+	0
	>3292807	/42959			
35	len =	1450	nex =	2	
	Term	18522	17812	-	0
	Init	19126	18607	-	0
40	>3292807	/13264			
	len =	2439	nex =	9	
45	Init	20030	20399	+	0
	Intr	20826	20938	+	0
	Intr	21037	21133	+	0
	Intr	21226	21326	+	0
	Intr	21422	21478	+	0
50	Intr	21563	21639	+	0
	Intr	21733	21895	+	0
	Intr	21975	22072	+	0
	Term	22302	22468	+	0
55	>3292807	/21391			
	len =	2384	nex =	9	
	Init	20085	20399	+	0
60	Intr	20826	20938	+	0

1150

5	Intr	21037	21133	+	0
	Intr	21226	21326	+	0
	Intr	21422	21478	+	0
	Intr	21563	21639	+	0
	Intr	21733	21895	+	0
	Intr	21975	22072	+	0
	Term	22302	22468	+	0
>3292807		/95912			
10	len =	347	nex =	1	
	Sngl	22302	22507	+	0
>3292807		/29391			
20	len =	1588	nex =	4	
	Term	30143	29734	-	0
	Intr	30387	30220	-	0
	Intr	30755	30582	-	0
	Init	31321	30873	-	0
>3292807		/8228			
25	len =	625	nex =	2	
	Term	73592	73242	-	0
	Init	73866	73675	-	0
>3292807		/25628			
35	len =	2590	nex =	6	
	Term	73592	73176	-	0
	Intr	73782	73675	-	0
	Intr	74304	74203	-	0
	Intr	74936	74735	-	0
40	Intr	75046	75018	-	0
	Init	75763	75408	-	0
>3293581		/13181			
45	len =	1475	nex =	3	
	Term	69408	68947	-	0
	Intr	69691	69496	-	0
	Init	70421	70034	-	0
>3293581		/1461			
55	len =	1570	nex =	3	
	Term	69408	68922	-	0
	Intr	69691	69496	-	0
	Init	70489	70034	-	0
>3293581		/39258			
60	len =	1590	nex =	3	

1151

	Term	69408	68907	-	0
	Intr	69691	69496	-	0
	Init	70496	70034	-	0
5	>3293582 /97031				
	len =	390	nex =	1	
10	Sngl	22292	21903	-	0
	>3293582 /14258				
	len =	3176	nex =	7	
15	Term	50181	49892	-	0
	Intr	50364	50280	-	0
	Intr	51074	50997	-	0
	Intr	51309	51162	-	0
20	Intr	51674	51515	-	0
	Intr	52551	52516	-	0
	Init	53067	53029	-	0
	>3293583 /111377				
25	len =	1211	nex =	1	
	Sngl	46281	46069	-	0
30	>3293583 /1463				
	len =	1510	nex =	4	
	Term	51608	51036	-	0
35	Intr	51864	51769	-	0
	Intr	52214	51938	-	0
	Init	52536	52350	-	0
	>3293583 /112432				
40	len =	1849	nex =	3	
	Term	55125	55072	-	0
	Intr	55543	55514	-	0
45	Init	55756	55620	-	0
	>3297806 /24629				
	len =	971	nex =	2	
50	Term	18135	18036	-	0
	Init	18729	18586	-	0
	>3297806 /21867				
55	len =	1979	nex =	7	
	Term	41785	41666	-	0
	Intr	41937	41872	-	0
60	Intr	42102	42007	-	0

1152

	Intr	42340	42238	-	0
	Intr	42580	42423	-	0
	Intr	42933	42811	-	0
	Init	43304	43034	-	0
5	>3297806 /2306				
	len =	1972	nex =	7	
10	Term	41785	41666	-	0
	Intr	41937	41872	-	0
	Intr	42102	42007	-	0
	Intr	42340	42238	-	0
	Intr	42580	42423	-	0
15	Intr	42933	42811	-	0
	Init	43304	43034	-	0
	>3297806 /40123				
20	len =	2093	nex =	4	
	Init	45275	45667	+	0
	Intr	45924	46300	+	0
	Intr	46640	46876	+	0
25	Term	46967	47367	+	0
	>3297806 /24360				
	len =	988	nex =	5	
30	Init	47770	47885	+	0
	Intr	47971	47992	+	0
	Intr	48108	48213	+	0
	Intr	48295	48402	+	0
35	Term	48496	48757	+	0
	>3297806 /14385				
	len =	1719	nex =	4	
40	Init	50596	50900	+	0
	Intr	51195	51391	+	0
	Intr	51826	51877	+	0
	Term	51988	52314	+	0
45	>3297806 /8161				
	len =	2257	nex =	4	
50	Term	67615	67335	-	0
	Intr	67811	67699	-	0
	Intr	68621	68463	-	0
	Init	69591	69324	-	0
55	>3297806 /9341				
	len =	790	nex =	2	
60	Term	72925	72662	-	0
	Init	73446	73018	-	0

>3297806 /6896

5	len =	2658	nex =	8	
	Term	83301	82800	-	0
	Intr	83518	83381	-	0
	Intr	83942	83742	-	0
	Intr	84248	84171	-	0
10	Intr	84531	84352	-	0
	Intr	84770	84621	-	0
	Intr	85083	84883	-	0
	Init	85457	85176	-	0

15 >3298532 /18983

	len =	1557	nex =	1	
20	Sngl	10434	8878	-	0

>3298532 /2114

	len =	1830	nex =	5	
25	Term	26390	26016	-	0
	Intr	26666	26475	-	0
	Intr	26876	26747	-	0
	Intr	27100	26958	-	0
	Init	27845	27756	-	0

30 >3298532 /630

	len =	2088	nex =	4	
35	Init	45179	45777	+	0
	Intr	45865	46206	+	0
	Intr	46287	46361	+	0
	Term	46938	47266	+	0

40 >3298532 /16625

	len =	1601	nex =	3	
	Init	48070	48390	+	0
45	Intr	49014	49037	+	0
	Term	49150	49670	+	0

>3298532 /28303

50	len =	2115	nex =	7	
	Term	54352	53910	-	0
	Intr	54548	54474	-	0
	Intr	54758	54684	-	0
55	Intr	54914	54849	-	0
	Intr	55095	55019	-	0
	Intr	55479	55310	-	0
	Init	56024	55735	-	0

60 >3298532 /41121

1154

	len =	454	nex =	1	
5	Sngl	64301	64754	+	0
	>3299824	/29869			
	len =	2330	nex =	5	
10	Term	16577	16462	-	0
	Intr	17378	17142	-	0
	Intr	17531	17470	-	0
	Intr	18024	17899	-	0
	Init	18270	18221	-	0
15	>3299824	/20753			
	len =	1870	nex =	5	
20	Init	18439	18529	+	0
	Intr	18632	18674	+	0
	Intr	19090	19131	+	0
	Intr	19213	19441	+	0
	Term	19805	20301	+	0
25	>3299824	/15442			
	len =	617	nex =	2	
30	Init	25851	25868	+	0
	Term	25952	26467	+	0
	>3299824	/711			
35	len =	1732	nex =	2	
	Init	35976	36226	+	0
	Term	37044	37707	+	0
40	>3299824	/113135			
	len =	1721	nex =	2	
45	Init	35987	36226	+	0
	Term	37044	37707	+	0
	>3299824	/17748			
	len =	1376	nex =	4	
50	Init	48724	48965	+	0
	Intr	49089	49189	+	0
	Intr	49289	49482	+	0
	Term	49567	50099	+	0
55	>3299824	/93014			
	len =	960	nex =	3	
60	Init	73060	73121	+	0

1155

	Intr	73213	73463	+	0
	Term	73906	74019	+	0
	>3299824 /111207				
5	len =	1604	nex =	3	
	Init	97176	97361	+	0
	Intr	97704	97837	+	0
10	Term	97933	98779	+	0
	>3299824 /37727				
	len =	1410	nex =	2	
15	Init	99110	99786	+	0
	Term	99950	100519	+	0
	>3309259 /40330				
20	len =	1966	nex =	3	
	Init	85348	85495	+	0
	Intr	85653	85944	+	0
25	Term	86040	87313	+	0
	>3309259 /27711				
	len =	1630	nex =	4	
30	Term	89111	88858	-	0
	Intr	89328	89210	-	0
	Intr	89667	89615	-	0
	Init	89802	89747	-	0
35	>3309259 /119712				
	len =	1705	nex =	4	
40	Term	89111	88841	-	0
	Intr	89328	89210	-	0
	Intr	89667	89615	-	0
	Init	89802	89747	-	0
	>3309259 /118778				
45	len =	987	nex =	1	
	Sngl	93145	92159	-	0
50	>3309276 /7149				
	len =	1455	nex =	6	
	Term	49451	49232	-	0
55	Intr	49635	49550	-	0
	Intr	49804	49734	-	0
	Intr	50152	50051	-	0
	Intr	50352	50254	-	0
60	Init	50686	50577	-	0

1156

>3319339 /6836

len = 1231 nex = 1

5 Sngl 1451 2681 + 0

>3319339 /34540

len = 1813 nex = 5

10 Init 45077 45206 + 0
 Intr 45305 45474 + 0
 Intr 45606 45885 + 0
 Intr 45999 46165 + 0
 15 Term 46268 46560 + 0

>3319339 /32443

len = 2033 nex = 5

20 Term 3121 2602 - 0
 Intr 3336 3204 - 0
 Intr 3716 3488 - 0
 Intr 4312 4280 - 0
 25 Init 4634 4405 - 0

>3319339 /1360

len = 1870 nex = 7

30 Init 52990 53088 + 0
 Intr 53171 53228 + 0
 Intr 53336 53462 + 0
 Intr 53548 53717 + 0
 35 Intr 53870 54149 + 0
 Intr 54243 54409 + 0
 Term 54518 54850 + 0

>3319339 /27916

len = 1832 nex = 7

Init 52992 53088 + 0
 Intr 53171 53228 + 0
 45 Intr 53336 53462 + 0
 Intr 53548 53717 + 0
 Intr 53870 54149 + 0
 Intr 54243 54409 + 0
 Term 54518 54823 + 0

>3319339 /27199

len = 2013 nex = 6

55 Init 70541 70591 + 0
 Intr 70680 70806 + 0
 Intr 70898 71067 + 0
 Intr 71347 71626 + 0
 Intr 71727 71875 + 0
 60 Term 71991 72346 + 0

202500-942500

1157

>3319339 /19839
 5 len = 1038 nex = 2
 Term 86289 85954 - 0
 Init 86991 86748 - 0
 >3319365 /26983
 10 len = 1471 nex = 5
 Term 25198 24914 - 0
 Intr 25328 25285 - 0
 15 Intr 25499 25442 - 0
 Intr 25917 25824 - 0
 Init 26384 26126 - 0
 >3319365 /113133
 20 len = 1059 nex = 1
 Sngl 39553 39787 + 0
 25 >3319365 /37317
 len = 472 nex = 1
 Sngl 39061 38590 - 0
 30 >3319365 /33945
 len = 370 nex = 1
 35 Sngl 46052 45690 - 0
 >3327922 /101361
 40 len = 938 nex = 2
 Term 25827 25600 - 0
 Init 26537 25924 - 0
 >3327922 /38439
 45 len = 2141 nex = 5
 Init 32746 32973 + 0
 Intr 33067 33167 + 0
 50 Intr 33246 33289 + 0
 Intr 33713 34348 + 0
 Term 34423 34886 + 0
 >3327922 /19349
 55 len = 591 nex = 1
 Sngl 35078 35668 + 0
 60 >3327922 /27727

1159

	len =	831	nex =	1	
5	Sngl	80171	81001	+	0
	>3327922	/34656			
	len =	801	nex =	1	
10	Sngl	80201	81001	+	0
	>3327922	/39583			
	len =	490	nex =	1	
15	Sngl	80576	81064	+	0
	>3327922	/13054			
20	len =	370	nex =	1	
	Sngl	80640	80995	+	0
	>3335331	/5586			
25	len =	793	nex =	3	
	Init	17573	17712	+	0
	Intr	17853	18139	+	0
30	Term	18220	18365	+	0
	>3335331	/32284			
	len =	4400	nex =	11	
35	Init	2322	2529	+	0
	Intr	2902	3011	+	0
	Intr	3108	3218	+	0
	Intr	3352	3416	+	0
40	Intr	3502	3583	+	0
	Intr	4342	4416	+	0
	Intr	4842	4912	+	0
	Intr	5033	5115	+	0
	Intr	5687	5751	+	0
45	Intr	5986	6060	+	0
	Term	6374	6721	+	0
	>3335331	/4326			
50	len =	953	nex =	2	
	Term	64184	63680	-	0
	Init	64632	64297	-	0
55	>3335331	/6906			
	len =	3250	nex =	8	
	Init	7810	8237	+	0
60	Intr	8419	8481	+	0

1160

	Intr	8573	8668	+	0
	Intr	9068	9115	+	0
	Intr	9290	9361	+	0
	Intr	10005	10063	+	0
5	Intr	10270	10360	+	0
	Term	10857	11050	+	0
	>3335331		/19247		
10	len =	3193	nex =	8	
	Init	7832	8237	+	0
	Intr	8419	8481	+	0
	Intr	8573	8668	+	0
15	Intr	9068	9115	+	0
	Intr	9290	9361	+	0
	Intr	10005	10063	+	0
	Intr	10270	10360	+	0
	Term	10857	11024	+	0
20	>3335331		/118658		
	len =	1125	nex =	2	
25	Term	79626	79511	-	0
	Init	80234	80142	-	0
	>3335331		/42267		
30	len =	939	nex =	2	
	Init	81876	82398	+	0
	Term	82623	82814	+	0
35	>3335331		/17230		
	len =	1210	nex =	1	
40	Sngl	83926	85126	+	0
	>3335356		/23238		
	len =	1810	nex =	2	
45	Term	99574	98995	-	0
	Init	100803	100401	-	0
	>3335356		/31586		
50	len =	4099	nex =	8	
	Init	40725	40991	+	0
	Intr	41490	41675	+	0
	Intr	41764	41876	+	0
55	Intr	41973	42156	+	0
	Intr	42435	42532	+	0
	Intr	43054	43261	+	0
	Intr	43348	43605	+	0
	Term	44286	44823	+	0
60					

1161

>3335356 /4723

len = 711 nex = 3

5	Init	48610	48831	+	0
	Intr	48899	48963	+	0
	Term	49052	49320	+	0

>3335356 /15751

10 len = 4072 nex = 10

	Term	57235	56932	-	0
	Intr	58046	57956	-	0
15	Intr	58325	58155	-	0
	Intr	58546	58430	-	0
	Intr	58799	58671	-	0
	Intr	59124	58900	-	0
	Intr	59544	59348	-	0
20	Intr	59678	59650	-	0
	Intr	60330	60263	-	0
	Init	61003	60621	-	0

>3335356 /7394

25 len = 1030 nex = 2

	Init	65862	66005	+	0
	Term	66091	66596	+	0

30

>3335356 /25934

len = 910 nex = 3

35	Init	67764	68104	+	0
	Intr	68204	68310	+	0
	Term	68403	68673	+	0

>3335356 /19998

40 len = 1039 nex = 1

Sngl 79183 80221 + 0

45 >3337347 /4598

len = 2715 nex = 7

	Init	32576	32711	+	0
50	Intr	33094	33465	+	0
	Intr	33623	33850	+	0
	Intr	33937	34291	+	0
	Intr	34378	34502	+	0
	Intr	34599	34734	+	0
55	Term	34820	35288	+	0

>3337347 /36037

60 len = 1937 nex = 6

1162

	Init	33121	33465	+	0
	Intr	33623	33850	+	0
	Intr	33937	34291	+	0
	Intr	34378	34502	+	0
5	Intr	34599	34734	+	0
	Term	34820	35057	+	0

>3337347 /37499

10	len =	2676	nex =	14	
	Init	36337	36480	+	0
	Intr	36582	36614	+	0
	Intr	36700	36764	+	0
15	Intr	36846	36933	+	0
	Intr	37031	37103	+	0
	Intr	37379	37485	+	0
	Intr	37592	37691	+	0
	Intr	37841	37917	+	0
20	Intr	38020	38068	+	0
	Intr	38160	38206	+	0
	Intr	38297	38367	+	0
	Intr	38476	38531	+	0
	Intr	38603	38732	+	0
25	Term	38833	39012	+	0

>3337347 /16411

30	len =	2129	nex =	3	
	Term	48472	48232	-	0
	Intr	49760	49222	-	0
	Init	50360	50200	-	0

35 >3337347 /29146

	len =	702	nex =	1	
40	Sngl	60015	59314	-	0

>3337347 /23727

	len =	979	nex =	1	
45	Sngl	95243	96221	+	0

>3341671 /105595

50	len =	1168	nex =	2	
	Term	32434	32064	-	0
	Init	32805	32715	-	0

>3341671 /39499

55	len =	2552	nex =	6	
	Term	32434	32032	-	0
	Intr	32805	32715	-	0
60	Intr	33301	33190	-	0

1163

	Intr	33563	33475	-	0
	Intr	33731	33661	-	0
	Init	34583	34277	-	0
5	>3341671	/5217			
	len =	610	nex =	2	
10	Init	39041	39309	+	0
	Term	39394	39647	+	0
	>3341671	/30696			
15	len =	133	nex =	1	
	Sngl	41075	40943	-	0
	>3341671	/40641			
20	len =	869	nex =	2	
	Init	45689	45897	+	0
	Term	46011	46557	+	0
25	>3341671	/111157			
	len =	610	nex =	2	
30	Init	45812	45897	+	0
	Term	46011	46420	+	0
	>3341671	/40953			
35	len =	1719	nex =	5	
	Term	47540	47148	-	0
	Intr	47874	47709	-	0
	Intr	48054	47945	-	0
	Intr	48358	48140	-	0
40	Init	48866	48453	-	0
	>3341671	/1136			
45	len =	314	nex =	1	
	Sngl	58089	57776	-	0
	>3341671	/42237			
50	len =	1709	nex =	4	
	Term	68153	67782	-	0
	Intr	68844	68674	-	0
	Intr	69172	69084	-	0
55	Init	69484	69280	-	0
	>3341671	/34828			
60	len =	1345	nex =	6	

2025 RELEASE UNDER E.O. 14176

1164

	Init	72023	72154	+	0
	Intr	72231	72327	+	0
	Intr	72444	72516	+	0
	Intr	72596	72697	+	0
5	Intr	72783	73032	+	0
	Term	73118	73367	+	0
	>3341671		/36996		
10	len =	3070	nex =	6	
	Init	78677	78786	+	0
	Intr	78936	79004	+	0
	Intr	79193	79308	+	0
15	Intr	79630	79750	+	0
	Intr	80100	80194	+	0
	Term	80294	80892	+	0
	>3341671		/114909		
20	len =	1518	nex =	6	
	Init	78682	78786	+	0
	Intr	78936	79004	+	0
	Intr	79193	79308	+	0
25	Intr	79376	79445	+	0
	Intr	79630	79750	+	0
	Term	80100	80194	+	0
	>3341671		/19760		
30	len =	1278	nex =	2	
	Term	83127	82506	-	0
35	Init	83783	83570	-	0
	>3355463		/11583		
40	len =	1192	nex =	4	
	Term	19567	19377	-	0
	Intr	19829	19673	-	0
	Intr	20349	20214	-	0
	Init	20568	20449	-	0
45					
	>3355463		/124576		
	len =	894	nex =	2	
50	Init	24459	24491	+	0
	Term	24870	25152	+	0
	>3355463		/22479		
55	len =	1198	nex =	2	
	Init	44058	44421	+	0
	Term	44510	44770	+	0
60	>3355463		/27485		

1165

	len =	1150	nex =	1	
	Sngl	43581	44721	+	0
5	>3355463	/16403			
	len =	1510	nex =	8	
10	Init	44997	45084	+	0
	Intr	45167	45248	+	0
	Intr	45334	45438	+	0
	Intr	45532	45605	+	0
	Intr	45692	45814	+	0
15	Intr	45902	45973	+	0
	Intr	46058	46154	+	0
	Term	46240	46506	+	0
	>3355463	/3850			
20	len =	1187	nex =	5	
	Term	60749	60655	-	0
	Intr	61079	60839	-	0
25	Intr	61271	61155	-	0
	Intr	61421	61359	-	0
	Init	61841	61597	-	0
	>3355463	/15933			
30	len =	2560	nex =	9	
	Term	90887	90794	-	0
	Intr	91034	90970	-	0
35	Intr	91211	91128	-	0
	Intr	91605	91540	-	0
	Intr	91892	91758	-	0
	Intr	92227	92182	-	0
	Intr	92490	92384	-	0
40	Intr	93059	93011	-	0
	Init	93353	93139	-	0
	>3366536	/27918			
45	len =	1725	nex =	6	
	Term	26503	26130	-	0
	Intr	26880	26613	-	0
	Intr	27116	26996	-	0
50	Intr	27259	27206	-	0
	Intr	27499	27394	-	0
	Init	27854	27748	-	0
	>3366536	/32438			
55	len =	755	nex =	4	
	Init	32527	32606	+	0
	Intr	32715	32774	+	0
60	Intr	32882	32935	+	0

1166

	Term	33014	33281	+	0
	>3366536	/5485			
5	len =	1553	nex =	2	
	Term	73632	73175	-	0
	Init	74390	73723	-	0
10	>3367500	/11257			
	len =	1890	nex =	2	
	Init	11335	11645	+	0
15	Term	11808	12594	+	0
	>3367500	/125642			
	len =	2198	nex =	4	
20	Init	27603	28220	+	0
	Intr	28298	28420	+	0
	Intr	28512	28628	+	0
	Term	29428	29800	+	0
25	>3367500	/21771			
	len =	2129	nex =	4	
30	Init	27671	28220	+	0
	Intr	28298	28420	+	0
	Intr	28512	28628	+	0
	Term	29428	29799	+	0
35	>3367500	/104934			
	len =	1930	nex =	4	
	Init	27736	28220	+	0
40	Intr	28298	28420	+	0
	Intr	28512	28628	+	0
	Term	29428	29665	+	0
	>3367500	/25284			
45	len =	865	nex =	3	
	Init	27765	28220	+	0
	Intr	28298	28420	+	0
50	Term	28512	28629	+	0
	>3367500	/37435			
	len =	2016	nex =	10	
55	Term	29954	29748	-	0
	Intr	30174	30040	-	0
	Intr	30390	30253	-	0
	Intr	30575	30469	-	0
60	Intr	30799	30669	-	0

2025 RELEASE UNDER E.O. 14176

1167

	Intr	30915	30875	-	0
	Intr	31085	30985	-	0
	Intr	31254	31167	-	0
	Intr	31505	31404	-	0
5	Init	31763	31585	-	0
	>3367500		/35575		
	len =	281	nex =	1	
10	Sngl	44549	44829	+	0
	>3367567		/95636		
15	len =	411	nex =	1	
	Sngl	1	411	+	0
	>3367567		/148676		
20	len =	1341	nex =	4	
	Term	16182	16130	-	0
	Intr	16440	16386	-	0
25	Intr	16623	16539	-	0
	Init	17470	16947	-	0
	>3367567		/16204		
30	len =	1178	nex =	6	
	Term	31280	31024	-	0
	Intr	31447	31367	-	0
	Intr	31568	31524	-	0
35	Intr	31726	31669	-	0
	Intr	31913	31827	-	0
	Init	32050	31996	-	0
	>3367567		/31322		
40	len =	1172	nex =	6	
	Term	31280	31030	-	0
	Intr	31447	31367	-	0
45	Intr	31568	31524	-	0
	Intr	31726	31669	-	0
	Intr	31913	31827	-	0
	Init	32050	31996	-	0
50	>3367567		/34272		
	len =	1514	nex =	1	
	Sngl	48006	49519	+	0
55	>3367567		/4745		
	len =	614	nex =	2	
60	Term	74205	73805	-	0

1168

Init	74418	74292	-	0
------	-------	-------	---	---

>3367567 /17117

```
5      len =      1216      nex =          3
```

Term	84942	84402	-	0
------	-------	-------	---	---

Intr	85322	85058	-	0
------	-------	-------	---	---

Init	85617	85409	-	0
------	-------	-------	---	---

10

>3367567 /42841

```
len = 2110    nex = 7
```

15	Term	86680	86417	-	0
----	------	-------	-------	---	---

Intr	86891	86754	-	0
------	-------	-------	---	---

Intr	87123	86976	-	0
------	-------	-------	---	---

Intr	87386	87274	-	0
------	-------	-------	---	---

Intr	87739	87483	-	0
------	-------	-------	---	---

20	Intr	88305	88020	-	0
----	------	-------	-------	---	---

Init	88518	88382	-	0
------	-------	-------	---	---

>3367567 /13699

```
25      len =      1019      nex =          4
```

Init	88880	89082	+	0
------	-------	-------	---	---

Intr	89175	89371	+	0
------	-------	-------	---	---

Intr	89462	89532	+	0
------	-------	-------	---	---

30	Term	89633	89686	+	0
----	------	-------	-------	---	---

>3386593 /41984

```
len =      562      nex =      1
```

35

Sngl	14092	14653	+	0
------	-------	-------	---	---

>3386593 / 82

```
40      len =      140      nex =      1
```

Snq1	16606	16467	-	0
------	-------	-------	---	---

>3386593 /19844

```

45         len = 1150     nex = 3

```

Init	4750	5187	+	0
------	------	------	---	---

Intr	5328	5456	+	0
------	------	------	---	---

50	Term	5544	5890	+	0
----	------	------	------	---	---

>3386593 /18854

```
len = 1713    nex = 4
```

55

Intr	63184	63237	+	0
------	-------	-------	---	---

Intr	63333	63420	+	0
------	-------	-------	---	---

Term	63493	63944	+	0
------	-------	-------	---	---

60

1169

>3386593 /95396

```
len = 1213    nex = 3
```

5	Init	62306	62493	+	0
	Intr	63184	63237	+	0
	Term	63333	63420	+	0

>3386593 /998

```
10      len =      2124      nex =      10
```

15	Term	69215	69011	-	0
	Intr	69358	69307	-	0
	Intr	69546	69457	-	0
	Intr	69727	69632	-	0
	Intr	69907	69808	-	0
20	Intr	70047	69992	-	0
	Intr	70209	70174	-	0
	Intr	70369	70293	-	0
	Intr	70834	70768	-	0
	Init	71134	70927	-	0

>3386593 /13925

```

25         len = 1676     nex = 3

```

	Init	7808	7937	+	0
	Intr	8704	9013	+	0
30	Term	9133	9483	+	0

>3386593 /12250

```
len = 1667 nex = 7
```

	Term	85697	85517	-	0
	Intr	86125	85952	-	0
	Intr	86285	86211	-	0
	Intr	86455	86375	-	0
40	Intr	86624	86550	-	0
	Intr	86781	86707	-	0
	Init	87183	86857	-	0

>3386593 /12487

```

45      len =      1630      nex =          6

```

	Term	85849	85656	-	0
	Intr	86125	85952	-	0
50	Intr	86285	86211	-	0
	Intr	86455	86375	-	0
	Intr	86624	86550	-	0
	Intr	86781	86707	-	0

>3386593 /9000

```
len = 2770 nex = 13
```

	Init	87778	87938	+	0
60	Intr	88044	88171	+	0

1170

	Intr	88293	88398	+	0
	Intr	88487	88568	+	0
	Intr	88661	88731	+	0
	Intr	88818	88859	+	0
5	Intr	88982	89106	+	0
	Intr	89196	89289	+	0
	Intr	89407	89482	+	0
	Intr	89599	89762	+	0
10	Intr	89840	89945	+	0
	Intr	90121	90207	+	0
	Term	90311	90531	+	0
>3395421		/123915			
15	len =	625	nex =	2	
	Init	14364	14582	+	0
	Term	14746	14988	+	0
20	>3395421	/14329			
	len =	879	nex =	3	
	Term	18618	18349	-	0
25	Intr	18781	18712	-	0
	Init	19227	18940	-	0
>3395421		/43031			
30	len =	1786	nex =	4	
	Term	40343	39798	-	0
	Intr	40877	40709	-	0
	Intr	41167	40976	-	0
35	Init	41583	41311	-	0
>3395421		/25615			
	len =	833	nex =	1	
40	Sngl	50675	51507	+	0
>3395421		/16674			
45	len =	1056	nex =	3	
	Init	9467	9820	+	0
	Intr	9908	10031	+	0
	Term	10113	10522	+	0
50	>3399678	/3807			
	len =	2443	nex =	6	
55	Term	9355	8974	-	0
	Intr	9578	9453	-	0
	Intr	9939	9691	-	0
	Intr	10435	10084	-	0
	Intr	10680	10541	-	0
60	Init	11416	10972	-	0

1171

>3399678 /39874

5	len =	2953	nex =	11	
	Init	11842	11976	+	0
	Intr	12277	12410	+	0
	Intr	12562	12643	+	0
	Intr	12772	12839	+	0
10	Intr	12966	13067	+	0
	Intr	13154	13243	+	0
	Intr	13590	13664	+	0
	Intr	13766	13798	+	0
	Intr	13901	13978	+	0
15	Intr	14081	14137	+	0
	Term	14395	14794	+	0

>3399678 /3234

```
20      len =      315      nex =      1
      Sngl 3776      4090      +      0
```

>3399678 /12077

25	len =	1815	nex =	5	
	Init	4154	4420	+	0
	Intr	4549	4621	+	0
30	Intr	5205	5269	+	0
	Intr	5617	5726	+	0
	Term	5816	5968	+	

>3399678 /6672

35	len =	3673	nex =	12	
	Init	53802	54056	+	0
	Intr	54661	54805	+	0
40	Intr	54887	54982	+	0
	Intr	55372	55448	+	0
	Intr	55709	55788	+	0
	Intr	55934	55994	+	0
	Intr	56137	56204	+	0
45	Intr	56296	56377	+	0
	Intr	56467	56556	+	0
	Intr	56645	56722	+	0
	Intr	56864	56983	+	0
	Term	57108	57474	+	0

>3399678 /38167

	len =	2082	nex =	9	
55	Init	55396	55448	+	0
	Intr	55709	55788	+	0
	Intr	55934	55994	+	0
	Intr	56137	56204	+	0
	Intr	56296	56377	+	0
60	Intr	56467	56556	+	0

1172

Intr	56645	56722	+	0
Intr	56864	56983	+	0
Term	57108	57474	+	0

5 >3399678 /108783

```
len = 111    nex = 1
```

Snq1	6456	6566	+	0
------	------	------	---	---

>3399678 /155960

```
len =      513      nex =      1
```

15	Snq1	73631	74143	+	0
----	------	-------	-------	---	---

>3399678 /13937

```
len = 1427    nex = 4
```

Init	84754	84947	+	0
Intr	85039	85251	+	0
Intr	85353	85508	+	0
Term	85609	86180	+	0

25
>3402671 /853

```
len = 3156    nex = 9
```

30	Term	104995	104737	-	0
	Intr	105259	105204	-	0
	Intr	105621	105565	-	0
	Intr	106216	106066	-	0
	Intr	106605	106557	-	0
35	Intr	106819	106726	-	0
	Intr	106986	106914	-	0
	Intr	107186	107086	-	0
	Init	107892	107260	-	0

40 >3402671 /39069

```
len = 2213    nex = 4
```

	Init	16497	16715	+	0
45	Intr	16919	17386	+	0
	Intr	17497	17952	+	0
	Term	18047	18451	+	0

>3402671 /33545

```
len = 2633    nex = 4
```

Init	21758	21965	+	0
Intr	22289	22750	+	0
Intr	23024	23482	+	0
Term	23572	23906	+	0

>3402671 /29616

```
60      len =      1257      nex =          6
```

1173

	Term	29445	29258	-	0
	Intr	29604	29548	-	0
	Intr	29812	29699	-	0
5	Intr	29980	29930	-	0
	Intr	30161	30064	-	0
	Init	30514	30442	-	0
	>3402695 /7469				
10	len =	1510	nex =	2	
	Term	13030	12578	-	0
	Init	13512	13440	-	0
15	>3402695 /152917				
	len =	653	nex =	0	
20	>3402695 /16715				
	len =	822	nex =	2	
	Term	44889	44516	-	0
25	Init	45337	45254	-	0
	>3402695 /123659				
	len =	945	nex =	3	
30	Term	49228	48980	-	0
	Intr	49706	49346	-	0
	Init	49924	49810	-	0
35	>3402695 /35600				
	len =	3715	nex =	12	
	Term	64962	64623	-	0
40	Intr	65129	65078	-	0
	Intr	65432	65316	-	0
	Intr	65870	65778	-	0
	Intr	66348	66129	-	0
	Intr	66853	66694	-	0
45	Intr	67168	67094	-	0
	Intr	67504	67317	-	0
	Intr	67667	67586	-	0
	Intr	67847	67766	-	0
	Intr	68048	68014	-	0
50	Init	68337	68152	-	0
	>3402695 /6727				
	len =	897	nex =	1	
55	Sngl	70904	71800	+	0
	>3402745 /33333				
60	len =	1510	nex =	3	

20250922 09:22:50

1174

	Term	24071	23112	-	0
	Intr	24287	24158	-	0
	Init	24617	24377	-	0
5	>3402745 /21213				
	len =	922	nex =	2	
10	Init	29567	29934	+	0
	Term	30280	30488	+	0
	>3402745 /33975				
15	len =	2309	nex =	7	
	Init	30883	31350	+	0
	Intr	31838	31908	+	0
	Intr	31988	32069	+	0
20	Intr	32169	32287	+	0
	Intr	32549	32626	+	0
	Intr	32714	32833	+	0
	Term	32923	33191	+	0
25	>3402745 /10221				
	len =	2063	nex =	4	
	Init	33374	33615	+	0
30	Intr	33835	34236	+	0
	Intr	34556	34970	+	0
	Term	35046	35436	+	0
	>3402745 /20924				
35	len =	2230	nex =	4	
	Init	33384	33615	+	0
	Intr	33835	34236	+	0
40	Intr	34556	34970	+	0
	Term	35046	35608	+	0
	>3402745 /36892				
45	len =	1308	nex =	3	
	Term	51257	50714	-	0
	Intr	51406	51332	-	0
	Init	52021	51488	-	0
50	>3402745 /36836				
	len =	2969	nex =	14	
55	Term	52493	52292	-	0
	Intr	52706	52584	-	0
	Intr	52859	52793	-	0
	Intr	53091	53007	-	0
	Intr	53395	53343	-	0
60	Intr	53658	53611	-	0

1175

	Intr	53884	53742	-	0
	Intr	54086	53980	-	0
	Intr	54266	54184	-	0
	Intr	54433	54360	-	0
5	Intr	54609	54529	-	0
	Intr	54759	54707	-	0
	Intr	54938	54876	-	0
	Init	55260	55060	-	0
10	>3402745	/42465			
	len =	1648	nex =	1	
15	Sngl	75185	74973	-	0
	>3402745	/35503			
	len =	1638	nex =	1	
20	Sngl	75185	74973	-	0
	>3402745	/31790			
25	len =	2014	nex =	4	
	Term	76187	75789	-	0
	Intr	76568	76278	-	0
	Intr	76777	76655	-	0
	Init	77133	76861	-	0
30	>3402745	/4621			
	len =	2027	nex =	5	
35	Init	80600	80931	+	0
	Intr	81288	81467	+	0
	Intr	81659	81740	+	0
	Intr	81819	82085	+	0
	Term	82176	82626	+	0
40	>3402745	/95385			
	len =	588	nex =	2	
45	Init	82005	82085	+	0
	Term	82176	82592	+	0
	>3402745	/17760			
50	len =	2350	nex =	9	
	Init	85225	85495	+	0
	Intr	85763	85943	+	0
	Intr	86039	86134	+	0
55	Intr	86229	86284	+	0
	Intr	86453	86536	+	0
	Intr	86690	86782	+	0
	Intr	86876	86943	+	0
	Intr	87122	87195	+	0
60	Term	87314	87572	+	0

>3402745 /40247

5	len =	2298	nex =	9	
	Init	90176	90278	+	0
	Intr	90538	90721	+	0
	Intr	90829	90939	+	0
	Intr	91060	91115	+	0
10	Intr	91287	91370	+	0
	Intr	91504	91543	+	0
	Intr	91646	91713	+	0
	Intr	91818	91891	+	0
	Term	91975	92274	+	0

>3402747 /5691

	len =	1330	nex =	3	
20	Init	2703	3023	+	0
	Intr	3396	3691	+	0
	Term	3790	4023	+	0

>3402747 /31930

25	len =	564	nex =	2	
	Init	3464	3691	+	0
	Term	3790	4016	+	0

>3406034 /37757

	len =	2240	nex =	10	
35	Term	2322	2126	-	0
	Intr	2505	2408	-	0
	Intr	2729	2583	-	0
	Intr	2889	2828	-	0
	Intr	3253	3163	-	0
40	Intr	3389	3336	-	0
	Intr	3580	3494	-	0
	Intr	3801	3673	-	0
	Intr	4116	4046	-	0
	Init	4365	4221	-	0

>3406034 /28528

	len =	775	nex =	1	
50	Sngl	44408	43634	-	0

>3406034 /102813

	len =	1630	nex =	3	
55	Init	982	1230	+	0
	Intr	1313	1442	+	0
	Term	1995	2193	+	0

>3406034 /35331

1178

	len =	1543	nex =	4	
	Term	49244	48780	-	0
	Intr	49472	49347	-	0
5	Intr	49668	49615	-	0
	Init	50322	49851	-	0
	>3420042	/33195			
10	len =	1210	nex =	3	
	Init	51312	51412	+	0
	Intr	51803	51916	+	0
	Term	52019	52117	+	0
15	>3420042	/34564			
	len =	1630	nex =	4	
20	Init	51312	51412	+	0
	Intr	51803	51916	+	0
	Intr	52019	52155	+	0
	Term	52279	52540	+	0
25	>3420042	/6397			
	len =	913	nex =	1	
	Sngl	58119	57207	-	0
30	>3420043	/38891			
	len =	3670	nex =	5	
35	Term	32310	31803	-	0
	Intr	33332	33174	-	0
	Intr	33637	33456	-	0
	Intr	34403	34232	-	0
40	Init	35464	35021	-	0
	>3420043	/35872			
	len =	1066	nex =	0	
45	>3420043	/22599			
	len =	1104	nex =	0	
	>3420043	/124634			
50	len =	2510	nex =	1	
	Sngl	41769	41941	+	0
55	>3420043	/36979			
	len =	3413	nex =	8	
	Init	56129	56549	+	0
60	Intr	57173	57262	+	0

1179

5	Intr	57809	57875	+	0
	Intr	58355	58447	+	0
	Intr	58535	58720	+	0
	Intr	58809	58889	+	0
	Intr	58968	59105	+	0
	Term	59204	59541	+	0
>3420043		/9257			
10	len =	593	nex =	1	
	Sngl	61934	61342	-	0
>3420043		/19211			
20	len =	1522	nex =	6	
	Init	70609	70725	+	0
	Intr	70916	70985	+	0
	Intr	71126	71173	+	0
	Intr	71279	71302	+	0
	Intr	71525	71707	+	0
Term	71791	72130	+	0	
>3420043		/206307			
30	len =	1090	nex =	5	
	Init	70617	70725	+	0
	Intr	70916	70985	+	0
	Intr	71126	71173	+	0
	Intr	71279	71302	+	0
	Term	71525	71706	+	0
>3420043		/150229			
40	len =	1511	nex =	6	
	Init	70617	70725	+	0
	Intr	70916	70985	+	0
	Intr	71126	71173	+	0
	Intr	71279	71302	+	0
	Intr	71525	71707	+	0
Term	71791	72127	+	0	
>3420043		/13487			
50	len =	1518	nex =	0	
	>3426033 /10624				
55	len =	2499	nex =	11	
	Init	10293	10440	+	0
	Intr	10665	10760	+	0
	Intr	10841	10909	+	0
	Intr	11001	11105	+	0
	Intr	11183	11239	+	0
60	Intr	11323	11523	+	0
	Intr	11684	11776	+	0

1180

	Intr	11858	11957	+	0
	Intr	12082	12181	+	0
	Intr	12426	12511	+	0
	Term	12597	12791	+	0
5	>3426033	/33835			
	len =	393	nex =	2	
10	Init	29182	29215	+	0
	Term	29293	29574	+	0
	>3426033	/37620			
15	len =	3550	nex =	11	
	Term	57690	57420	-	0
	Intr	57870	57778	-	0
20	Intr	58073	58007	-	0
	Intr	58268	58163	-	0
	Intr	58579	58495	-	0
	Intr	59100	59009	-	0
	Intr	59365	59243	-	0
25	Intr	59760	59655	-	0
	Intr	60149	60043	-	0
	Intr	60338	60247	-	0
	Init	60965	60797	-	0
	>3445196	/33511			
30	len =	1643	nex =	4	
	Init	35160	35638	+	0
	Intr	35732	35801	+	0
35	Intr	36165	36327	+	0
	Term	36415	36802	+	0
	>3445196	/121762			
40	len =	854	nex =	0	
	>3445196	/39568			
	len =	3654	nex =	12	
45	Init	70556	70702	+	0
	Intr	70777	70922	+	0
	Intr	71023	71099	+	0
	Intr	71180	71302	+	0
50	Intr	71707	71809	+	0
	Intr	71878	71915	+	0
	Intr	72121	72219	+	0
	Intr	72321	72407	+	0
	Intr	72500	72568	+	0
55	Intr	72754	72825	+	0
	Intr	72906	73027	+	0
	Term	73847	74209	+	0
	>3445196	/35			
60					

1181

	len =	2265	nex =	6	
	Init	72121	72219	+	0
	Intr	72321	72407	+	0
5	Intr	72500	72568	+	0
	Intr	72754	72825	+	0
	Intr	72906	73027	+	0
	Term	73847	74140	+	0
10	>3445196 /3181				
	len =	558	nex =	2	
	Init	82071	82214	+	0
15	Term	82390	82431	+	0
	>3449311 /5361				
	len =	1236	nex =	2	
20	Term	27604	27255	-	0
	Init	28490	28213	-	0
	>3449311 /141753				
25	len =	1072	nex =	4	
	Term	47195	46888	-	0
	Intr	47404	47285	-	0
30	Intr	47631	47489	-	0
	Init	47959	47723	-	0
	>3449311 /30287				
35	len =	1117	nex =	4	
	Term	47195	46889	-	0
	Intr	47404	47285	-	0
	Intr	47631	47489	-	0
40	Init	48005	47723	-	0
	>3449311 /121388				
	len =	2530	nex =	5	
45	Init	51403	51752	+	0
	Intr	51838	51965	+	0
	Intr	52072	52305	+	0
	Intr	53291	53537	+	0
50	Term	53646	53926	+	0
	>3449311 /21192				
	len =	2530	nex =	5	
55	Init	51403	51752	+	0
	Intr	51838	51965	+	0
	Intr	52072	52305	+	0
	Intr	53291	53537	+	0
60	Term	53646	53931	+	0

2025 RELEASE UNDER E.O. 14176

1182

>3449312 /114786

5	len =	1491	nex =	6	
	Init	20409	20522	+	0
	Intr	20834	20899	+	0
	Intr	21007	21151	+	0
	Intr	21246	21464	+	0
10	Intr	21559	21772	+	0
	Term	21862	21890	+	0

>3449312 /20725

15	len =	2155	nex =	7	
	Init	35173	35268	+	0
	Intr	35340	35570	+	0
20	Intr	35843	35940	+	0
	Intr	36031	36217	+	0
	Intr	36303	36433	+	0
	Intr	36515	36751	+	0
	Term	36841	37327	+	0

25 >3449312 /2159

```
len =      881    nex =      1
```

Sngl	40080	40960	+	0
------	-------	-------	---	---

>3449312 /11266

```
len =      2506      nex =          4
```

35	Init	70384	70926	+	0
	Intr	71299	71633	+	0
	Intr	71731	71930	+	0
	Term	72582	72889	+	0

40 >3449312 /37540

```
len =      2540      nex =          8
```

45	Init	8228	8379	+	0
	Intr	8897	8980	+	0
	Intr	9081	9301	+	0
	Intr	9387	9465	+	0
	Intr	9654	9823	+	0
50	Intr	9965	10151	+	0
	Intr	10243	10320	+	0
	Term	10403	10767	+	0

>3449313 /40718

```
55     len =      743     nex =      2
```

Term	17412	16864	-	0
------	-------	-------	---	---

Init	17606	17499	-	0
------	-------	-------	---	---

60 >3449313 /6699

1183

	len =	2291	nex =	3	
5	Term	19583	18826	-	0
	Intr	20255	19887	-	0
	Init	21116	20331	-	0
	>3449313 /31573				
10	len =	1964	nex =	3	
	Init	27831	28026	+	0
	Intr	28183	28618	+	0
	Term	29403	29794	+	0
15	>3449313 /147765				
	len =	831	nex =	1	
20	Sngl	55558	54728	-	0
	>3449314 /25475				
25	len =	698	nex =	1	
	Sngl	1344	647	-	0
	>3449314 /6982				
30	len =	1359	nex =	2	
	Term	6894	6520	-	0
	Init	7878	7794	-	0
35	>3449315 /29551				
	len =	1560	nex =	5	
40	Term	2958	2620	-	0
	Intr	3194	3059	-	0
	Intr	3363	3280	-	0
	Intr	3509	3452	-	0
	Init	4179	3608	-	0
45	>3449316 /107101				
	len =	391	nex =	1	
50	Sngl	46435	46825	+	0
	>3449316 /17050				
	len =	477	nex =	1	
55	Sngl	8487	8751	+	0
	>3449317 /3729				
60	len =	2698	nex =	9	

1184

	Term	36464	36043	-	0
	Intr	36675	36563	-	0
	Intr	36858	36770	-	0
	Intr	37059	36954	-	0
5	Intr	37343	37287	-	0
	Intr	37908	37819	-	0
	Intr	38239	38014	-	0
	Intr	38401	38368	-	0
	Init	38713	38596	-	0
10	>3449317 /2842				
	len =	1664	nex =	3	
15	Init	38942	39208	+	0
	Intr	39996	40060	+	0
	Term	40346	40605	+	0
20	>3449317 /91885				
	len =	643	nex =	3	
	Init	5652	5704	+	0
	Intr	5785	5868	+	0
25	Term	5959	6289	+	0
30	>3449317 /21756				
	len =	2140	nex =	7	
	Init	6801	6845	+	0
	Intr	7257	7326	+	0
	Intr	7438	7656	+	0
	Intr	7745	7824	+	0
35	Intr	7917	8033	+	0
	Intr	8150	8454	+	0
	Term	8552	8940	+	0
40	>3449318 /1523				
	len =	931	nex =	3	
	Term	21868	21538	-	0
	Intr	22060	21983	-	0
45	Init	22468	22163	-	0
50	>3449320 /31551				
	len =	954	nex =	3	
	Init	22442	22860	+	0
	Intr	23003	23177	+	0
	Term	23304	23395	+	0
55	>3449320 /14044				
	len =	1603	nex =	4	
	Init	37144	37440	+	0
60	Intr	38083	38151	+	0

1185

	Intr	38254	38360	+	0
	Term	38471	38746	+	0
	>3449320		/27973		
5	len =	1526	nex =	4	
	Init	48903	49084	+	0
	Intr	49585	49623	+	0
10	Intr	49797	50027	+	0
	Term	50116	50428	+	0
	>3449320		/12322		
15	len =	524	nex =	1	
	Sngl	60298	59775	-	0
	>3449320		/4875		
20	len =	984	nex =	2	
	Init	62405	62653	+	0
	Term	63001	63388	+	0
25	>3449321		/32212		
	len =	638	nex =	1	
30	Sngl	19476	20102	+	0
	>3449321		/6441		
35	len =	2658	nex =	2	
	Term	33609	32653	-	0
	Init	35310	34793	-	0
	>3449321		/27795		
40	len =	850	nex =	2	
	Term	36545	36330	-	0
	Init	37179	36861	-	0
45	>3449321		/14414		
	len =	1495	nex =	4	
50	Init	43130	43391	+	0
	Intr	43495	43751	+	0
	Intr	43847	43993	+	0
	Term	44204	44624	+	0
55	>3449322		/149970		
	len =	1056	nex =	2	
60	Term	15835	15245	-	0
	Init	16155	15991	-	0

1186

>3449322 /970

5	len =	496	nex =	1	
	Sngl	994	499	-	0

>3449323 /35890

10	len =	1873	nex =	3	
	Init	27590	27970	+	0
	Intr	28290	28562	+	0
	Term	28595	28757	+	0

15 >3449323 /17159

20	len =	1630	nex =	2	
	Init	32944	33507	+	0
	Term	34103	34565	+	0

>3449323 /21382

25	len =	276	nex =	1	
	Sngl	34337	34612	+	0

>3449323 /42747

30	len =	2074	nex =	4	
	Term	35032	34594	-	0
	Intr	35437	35219	-	0
35	Intr	35698	35519	-	0
	Init	36667	36397	-	0

>3449323 /39065

40	len =	869	nex =	0	
----	-------	-----	-------	---	--

>3449323 /13801

45	len =	2151	nex =	7	
	Term	52032	51755	-	0
	Intr	52225	52124	-	0
	Intr	52436	52311	-	0
	Intr	52789	52529	-	0
50	Intr	53298	52875	-	0
	Intr	53623	53505	-	0
	Init	53905	53724	-	0

>3449323 /122618

55	len =	1210	nex =	3	
	Term	69206	68963	-	0
	Intr	69709	69545	-	0
60	Init	69969	69811	-	0

20090909 14:59:59

>3449323 /13730

5	len =	1956	nex =	2	
	Term	71264	70606	-	0
	Init	71787	71394	-	0

>3449324 /19092

10	len =	710	nex =	3	
	Init	22002	22115	+	0
	Intr	22241	22398	+	0
15	Term	22512	22711	+	0

>3449325 /21890

20	len =	1523	nex =	3	
	Term	34360	33685	-	0
	Intr	34566	34431	-	0
	Init	35207	34772	-	0

>3449325 /42965

25	len =	1538	nex =	3	
	Term	34360	33734	-	0
30	Intr	34566	34431	-	0
	Init	35271	34772	-	0

>3449325 /13806

35	len =	1256	nex =	3	
	Term	40783	40075	-	0
	Intr	41008	40873	-	0
	Init	41330	41098	-	0

>3449326 /119717

40	len =	656	nex =	2	
45	Term	23133	22788	-	0
	Init	23443	23373	-	0

>3449326 /3086

50	len =	1090	nex =	2	
	Init	46084	46198	+	0
	Term	46486	47171	+	0

>3449326 /42926

55	len =	2685	nex =	9	
	Term	61606	61430	-	0
60	Intr	61803	61717	-	0

1188

	Intr	61962	61888	-	0
	Intr	62569	62465	-	0
	Intr	62770	62672	-	0
	Intr	62992	62859	-	0
5	Intr	63168	63090	-	0
	Intr	63494	63276	-	0
	Init	64114	63864	-	0
	>3449326 /120947				
10	len =	1295	nex =	2	
	Init	71680	72166	+	0
	Term	72722	72974	+	0
15	>3449327 /26543				
	len =	395	nex =	1	
20	Sngl	14331	13937	-	0
	>3449327 /37055				
25	len =	2110	nex =	10	
	Term	20457	20120	-	0
	Intr	20670	20567	-	0
	Intr	20855	20756	-	0
	Intr	21036	20939	-	0
30	Intr	21174	21122	-	0
	Intr	21586	21515	-	0
	Intr	21744	21673	-	0
	Intr	21920	21825	-	0
	Intr	22057	21990	-	0
35	Init	22225	22134	-	0
	>3449327 /27492				
40	len =	1820	nex =	4	
	Init	26052	26288	+	0
	Intr	26367	26724	+	0
	Intr	26809	27384	+	0
	Term	27475	27871	+	0
45	>3449327 /11250				
	len =	1693	nex =	1	
50	Sngl	35338	34719	-	0
	>3449327 /157644				
55	len =	1090	nex =	3	
	Term	5305	5089	-	0
	Intr	5612	5397	-	0
	Init	5960	5886	-	0
60	>3449327 /14401				

20250910.091600

1189

	len =	1180	nex =	3	
5	Term	5305	5037	-	0
	Intr	5612	5397	-	0
	Init	5960	5886	-	0
	>3449327 /38345				
10	len =	1496	nex =	3	
	Term	61922	61607	-	0
	Intr	62622	62493	-	0
	Init	63102	62707	-	0
15	>3449327 /12989				
	len =	1634	nex =	1	
20	Sngl	65098	64162	-	0
	>3449327 /39633				
25	len =	1858	nex =	3	
	Init	65970	66124	+	0
	Intr	66222	66328	+	0
	Term	67542	67827	+	0
30	>3449327 /3226				
	len =	1630	nex =	3	
35	Init	7695	7813	+	0
	Intr	7943	8038	+	0
	Term	8118	8562	+	0
	>3449327 /14416				
40	len =	1425	nex =	3	
	Init	7695	7813	+	0
	Intr	7943	8038	+	0
	Term	8118	8560	+	0
45	>3449329 /99348				
	len =	771	nex =	1	
50	Sngl	22128	21358	-	0
	>3449329 /39355				
	len =	456	nex =	1	
55	Sngl	22146	21709	-	0
	>3449329 /34273				
60	len =	1101	nex =	1	

1190

	Sngl	3712	2612	-	0
	>3449329 /33707				
5	len =	1699	nex =	5	
	Init	47513	47739	+	0
	Intr	47826	48010	+	0
10	Intr	48122	48322	+	0
	Intr	48410	48571	+	0
	Term	48642	49211	+	0
	>3449329 /117597				
15	len =	1750	nex =	1	
	Sngl	47513	47739	+	0
20	>3449329 /112975				
	len =	656	nex =	1	
	Sngl	77611	76956	-	0
25	>3449329 /7878				
	len =	2318	nex =	3	
30	Init	79072	79208	+	0
	Intr	79974	80125	+	0
	Term	80277	81389	+	0
	>3449329 /110801				
35	len =	575	nex =	2	
	Term	83763	83571	-	0
	Init	84145	84015	-	0
40	>3449330 /11513				
	len =	1658	nex =	3	
45	Term	63037	62811	-	0
	Intr	63717	63115	-	0
	Init	64468	64216	-	0
	>3449331 /18210				
50	len =	1246	nex =	3	
	Term	14046	13722	-	0
	Intr	14527	14422	-	0
55	Init	14967	14798	-	0
	>3449331 /35596				
60	len =	1469	nex =	2	

1191

Init	16127	16294	+	0
Term	17028	17595	+	0

5 >3449331 /6591

len = 2293 nex = 2

Init	18665	19043	+	0
Term	19547	20106	+	0

10 >3449331 /31901

len = 2110 nex = 11

15 Term 55268 55105 - 0

Intr 55402 55362 - 0

Intr 55630 55500 - 0

Intr 55809 55716 - 0

Intr 55980 55903 - 0

20 Intr 56162 56074 - 0

Intr 56315 56251 - 0

Intr 56488 56397 - 0

Intr 56676 56589 - 0

25 Intr 56832 56746 - 0

Init 57213 57145 - 0

>3449331 /7901

len = 2453 nex = 11

30 Term 55268 54818 - 0

Intr 55402 55362 - 0

Intr 55630 55500 - 0

Intr 55809 55716 - 0

35 Intr 55980 55903 - 0

Intr 56162 56074 - 0

Intr 56315 56251 - 0

Intr 56488 56397 - 0

40 Intr 56676 56589 - 0

Intr 56832 56746 - 0

Init 57270 57145 - 0

>3449331 /6701

45 len = 2621 nex = 11

Term 55268 54844 - 0

Intr 55402 55362 - 0

Intr 55630 55500 - 0

50 Intr 55809 55716 - 0

Intr 55980 55903 - 0

Intr 56162 56074 - 0

Intr 56315 56251 - 0

55 Intr 56488 56397 - 0

Intr 56676 56589 - 0

Intr 56832 56746 - 0

Init 57464 57342 - 0

>3449331 /24060

60

1192

	len =	1669	nex =	5	
	Term	59217	58869	-	0
	Intr	59381	59304	-	0
5	Intr	59645	59545	-	0
	Intr	59840	59743	-	0
	Init	60337	60235	-	0
	>3449331	/19444			
10	len =	701	nex =	2	
	Init	62474	62588	+	0
	Term	62729	62796	+	0
15	>3449331	/43057			
	len =	653	nex =	2	
20	Term	64577	64471	-	0
	Init	65123	64716	-	0
	>3449331	/7828			
25	len =	713	nex =	2	
	Init	68243	68660	+	0
	Term	68755	68955	+	0
30	>3449331	/16326			
	len =	702	nex =	2	
	Init	68246	68660	+	0
35	Term	68755	68947	+	0
	>3449331	/28773			
	len =	699	nex =	2	
40	Init	68251	68660	+	0
	Term	68755	68949	+	0
	>3449331	/94898			
45	len =	310	nex =	1	
	Sngl	68272	68572	+	0
50	>3449332	/31894			
	len =	1677	nex =	0	
	>3449332	/37326			
55	len =	2911	nex =	13	
	Term	30818	30775	-	0
	Intr	31080	30961	-	0
60	Intr	31445	31341	-	0

2025 RELEASE UNDER E.O. 14176

1193

	Intr	31611	31534	-	0
	Intr	31753	31688	-	0
	Intr	31902	31828	-	0
5	Intr	32046	31984	-	0
	Intr	32231	32142	-	0
	Intr	32404	32308	-	0
	Intr	32674	32577	-	0
	Intr	32859	32748	-	0
10	Intr	32971	32946	-	0
	Init	33267	33058	-	0
	>3449333	/37806			
15	len =	1510	nex =	1	
	Sngl	32409	33912	+	0
	>3449333	/39990			
20	len =	1218	nex =	1	
	Sngl	35415	34198	-	0
	>3449334	/37422			
25	len =	1129	nex =	3	
	Term	10778	10442	-	0
30	Intr	10889	10867	-	0
	Init	11570	11474	-	0
	>3449334	/11092			
35	len =	432	nex =	1	
	Sngl	38534	38322	-	0
	>3449334	/21208			
40	len =	1390	nex =	6	
	Init	48947	49088	+	0
	Intr	49267	49375	+	0
45	Intr	49516	49593	+	0
	Intr	49742	49775	+	0
	Intr	49871	49939	+	0
	Term	50057	50336	+	0
	>3449334	/12613			
50	len =	1460	nex =	3	
	Init	59084	59287	+	0
55	Intr	59884	60153	+	0
	Term	60375	60543	+	0
	>3449334	/9267			
60	len =	550	nex =	2	

1194

	Init	6629	6820	+	0
	Term	6919	7170	+	0
	>3449334 /6477				
5	len =	1114	nex =	2	
	Init	75520	75670	+	0
	Term	76364	76633	+	0
10	>3449334 /22052				
	len =	503	nex =	2	
15	Init	77334	77521	+	0
	Term	77615	77836	+	0
	>3449334 /154063				
20	len =	1339	nex =	5	
	Term	78133	77804	-	0
	Intr	78297	78209	-	0
	Intr	78506	78373	-	0
25	Intr	78651	78601	-	0
	Init	79142	78752	-	0
	>3449334 /4763				
30	len =	1158	nex =	2	
	Init	80747	81302	+	0
	Term	81744	81782	+	0
35	>3451055 /3982				
	len =	1589	nex =	4	
	Init	32795	33172	+	0
40	Intr	33446	33741	+	0
	Intr	33836	33976	+	0
	Term	34110	34383	+	0
	>3451055 /4578				
45	len =	2544	nex =	8	
	Init	41121	41326	+	0
	Intr	42072	42303	+	0
50	Intr	42380	42459	+	0
	Intr	42577	42730	+	0
	Intr	42805	42883	+	0
	Intr	42968	43055	+	0
	Intr	43172	43218	+	0
55	Term	43361	43664	+	0
	>3451055 /32181				
60	len =	2451	nex =	8	

1195

	Init	41200	41326	+	0
	Intr	42072	42303	+	0
	Intr	42380	42459	+	0
	Intr	42607	42730	+	0
5	Intr	42805	42883	+	0
	Intr	42968	43055	+	0
	Intr	43172	43218	+	0
	Term	43361	43650	+	0

10 >3451055 /25694

```
len = 2230    nex = 8
```

15	Init	61121	61435	+	0
	Intr	61520	61541	+	0
	Intr	61868	61924	+	0
	Intr	62011	62092	+	0
	Intr	62316	62422	+	0
20	Intr	62528	62652	+	0
	Intr	62751	62769	+	0
	Term	62860	63348	+	0

>3451055 /104784

```

25      len =      730      nex =      2
      Term  70038      69729      -      0
      Init  70454      70123      -      0

```

30 >3451055 /9836

```
len =      506    nex =      1
```

Snq1	78727	79232	+	0
------	-------	-------	---	---

>3461810 /11805

```
len = 2369    nex = 8
```

40	Term	42733	42323	-	0
	Intr	42938	42849	-	0
	Intr	43131	43045	-	0
	Intr	43376	43226	-	0
	Intr	43531	43464	-	0
45	Intr	43911	43840	-	0
	Intr	44222	44021	-	0
	Intr	44691	44486	-	0

>3461810 /27242

```

50      len =      1289      nex =      4

```

	Init	53744	54065	+	0
	Intr	54142	54297	+	0
55	Intr	54403	54601	+	0
	Term	54682	55032	+	0

>3461810 /10750

```
60      len =      1243      nex =          2
```

1196

	Term	7067	6782	-	0
	Init	8024	7769	-	0
5	>3461810 /2881				
	len =	1758	nex =	3	
	Term	97905	97123	-	0
10	Intr	98453	98141	-	0
	Init	98676	98543	-	0
	>3461834 /17914				
15	len =	1906	nex =	4	
	Term	11675	11426	-	0
	Intr	11885	11695	-	0
	Intr	12878	12560	-	0
20	Init	13331	13074	-	0
	>3461834 /38091				
	len =	472	nex =	1	
25	Sngl	29436	28965	-	0
	>3461834 /10164				
30	len =	1913	nex =	4	
	Init	49465	49733	+	0
	Intr	50347	50450	+	0
	Intr	50887	51016	+	0
35	Term	51126	51377	+	0
	>3461834 /30975				
	len =	2230	nex =	6	
40	Init	57851	58312	+	0
	Intr	58588	58751	+	0
	Intr	58845	59008	+	0
	Intr	59093	59352	+	0
45	Intr	59440	59566	+	0
	Term	59735	60074	+	0
	>3482964 /39928				
50	len =	2424	nex =	7	
	Term	23256	22843	-	0
	Intr	23516	23396	-	0
	Intr	23815	23590	-	0
55	Intr	23978	23891	-	0
	Intr	24191	24068	-	0
	Intr	24802	24731	-	0
	Init	24971	24885	-	0
60	>3482964 /40987				

1197

	len =	1881	nex =	6	
5	Init	27474	27850	+	0
	Intr	27982	28122	+	0
	Intr	28285	28425	+	0
	Intr	28664	28713	+	0
	Intr	28805	28934	+	0
10	Term	29067	29354	+	0
	>3482964	/31673			
	len =	263	nex =	1	
15	Sngl	5556	5818	+	0
	>3482964	/14105			
	len =	1707	nex =	3	
20	Init	77290	77640	+	0
	Intr	78070	78357	+	0
	Term	78446	78996	+	0
25	>3482964	/3709			
	len =	1403	nex =	6	
30	Term	7020	6611	-	0
	Intr	7158	7104	-	0
	Intr	7297	7245	-	0
	Intr	7457	7381	-	0
	Intr	7801	7552	-	0
35	Init	8013	7897	-	0
	>3482964	/34611			
	len =	1611	nex =	7	
40	Term	7020	6595	-	0
	Intr	7158	7104	-	0
	Intr	7297	7245	-	0
	Intr	7457	7381	-	0
	Intr	7801	7552	-	0
45	Intr	8050	7897	-	0
	Init	8205	8138	-	0
	>3482964	/11428			
	len =	2668	nex =	8	
50	Term	7020	6610	-	0
	Intr	7158	7104	-	0
	Intr	7297	7245	-	0
	Intr	7457	7381	-	0
	Intr	7801	7552	-	0
55	Intr	8050	7897	-	0
	Intr	8246	8138	-	0
	Init	8481	8418	-	0
60					

1198

>3492855 /38337

```
len =      4990      nex =      16
```

5	Init	25848	25975	+	0
	Intr	26062	26140	+	0
	Intr	26270	26331	+	0
	Intr	26602	26716	+	0
	Intr	26811	26886	+	0
10	Intr	26966	27061	+	0
	Intr	27278	27322	+	0
	Intr	27671	27727	+	0
	Intr	27827	27982	+	0
	Intr	28080	28145	+	0
15	Intr	28647	28770	+	0
	Intr	29541	29654	+	0
	Intr	29769	29846	+	0
	Intr	30008	30055	+	0
	Intr	30266	30334	+	0
20	Term	30495	30835	+	0

>3492855 /15984

```
len =      850    nex =      0
```

>3492855 /31772

```
len =      380      nex =      1
```

30	Snq1	67588	67967	+	0
----	------	-------	-------	---	---

>3492855 /42247

```
len = 502    nex = 1
```

Snq1	67588	68089	+	0
------	-------	-------	---	---

>3492855 /94363

```
40      len =      388      nex =      1
```

Sngl	67594	67981	+	0
------	-------	-------	---	---

>3492855 /12727

```
len = 698    nex = 1
```

Spa1 67594 68291 + 0

50 >3492855 /5339

```
len =      262      nex =      1
```

Sngl	67599	67860	+	0
------	-------	-------	---	---

55 >3492855 /8572

```
len =      337    nex =      1    1
```

60 Sngl 67621 67957 + 0

1199

```

>3492855      /34643
5      len =      692  nex =      1
      Sngl  67621  68312      +      0

>3492855      /8142
10     len =      698  nex =      1
      Sngl  67639  68336      +      0

>3492855      /113357
15     len =      351  nex =      1
      Sngl  67958  68308      +      0

20     >3510247      /105240
      len =      670  nex =      1
      Sngl  3626   4289      +      0

25     >3510247      /142022
      len =      226  nex =      1
      Sngl  39517  39742      +      0

30     >3510247      /26967
      len =      1512 nex =      6
35     Init  42812  43093      +      0
      Intr  43184  43305      +      0
      Intr  43393  43479      +      0
      Intr  43565  43627      +      0
40     Intr  43886  43923      +      0
      Term  44021  44323      +      0

>3510247      /26134
45     len =      574  nex =      3
      Init  42884  43093      +      0
      Intr  43184  43305      +      0
      Term  43393  43457      +      0

50     >3510336      /38743
      len =      2275 nex =      7
55     Term  10758  10440      -      0
      Intr  11079  10844      -      0
      Intr  11501  11396      -      0
      Intr  11676  11590      -      0
      Intr  12031  11813      -      0
60     Intr  12289  12119      -      0

```

2025 RELEASE UNDER E.O. 14176

	Init	12714	12390	-	1200 0
	>3510336	/29476			
5	len =	610	nex =	2	
	Init	23119	23477	+	0
	Term	23566	23719	+	0
10	>3510336	/18108			
	len =	762	nex =	2	
	Term	40411	40034	-	0
15	Init	40795	40541	-	0
	>3510337	/21311			
	len =	3939	nex =	6	
20	Init	1201	1383	+	0
	Intr	1756	1821	+	0
	Intr	3560	4202	+	0
	Intr	4287	4390	+	0
25	Intr	4470	4608	+	0
	Term	4752	5139	+	0
	>3510338	/26524			
30	len =	686	nex =	1	
	Sngl	7586	8271	+	0
	>3510339	/36971			
35	len =	2077	nex =	7	
	Term	13327	13214	-	0
	Intr	13556	13446	-	0
40	Intr	13746	13663	-	0
	Intr	13906	13838	-	0
	Intr	14081	14016	-	0
	Intr	14316	14191	-	0
	Init	14804	14412	-	0
45	>3510339	/26265			
	len =	310	nex =	1	
50	Sngl	18302	17998	-	0
	>3510339	/3732			
	len =	832	nex =	1	
55	Sngl	18846	18015	-	0
	>3510339	/13162			
60	len =	1665	nex =	4	

2025 RELEASE UNDER E.O. 14176

1201

	Init	25257	25329	+	0
	Intr	25742	25813	+	0
	Intr	26485	26565	+	0
5	Term	26649	26921	+	0
>3510339 /101308					
10	len =	1002	nex =	3	
	Init	27964	28382	+	0
	Intr	28442	28624	+	0
	Term	28735	28965	+	0
15	>3510339 /36973				
	len =	4244	nex =	17	
20	Init	32101	32339	+	0
	Intr	32424	32541	+	0
	Intr	32620	32769	+	0
	Intr	33398	33508	+	0
	Intr	33677	33760	+	0
25	Intr	34028	34126	+	0
	Intr	34225	34275	+	0
	Intr	34363	34412	+	0
	Intr	34484	34565	+	0
	Intr	34705	34799	+	0
30	Intr	34944	35010	+	0
	Intr	35123	35173	+	0
	Intr	35275	35370	+	0
	Intr	35459	35557	+	0
	Intr	35654	35683	+	0
35	Intr	35780	35845	+	0
	Term	36039	36344	+	0
>3510339 /6182					
40	len =	1225	nex =	6	
	Init	35127	35173	+	0
	Intr	35275	35370	+	0
	Intr	35459	35557	+	0
	Intr	35654	35683	+	0
45	Intr	35780	35845	+	0
	Term	36039	36351	+	0
>3510339 /7341					
50	len =	1630	nex =	5	
	Term	2222	1970	-	0
	Intr	2399	2316	-	0
	Intr	2651	2492	-	0
55	Intr	3073	3025	-	0
	Init	3595	3310	-	0
>3510339 /37148					
60	len =	1831	nex =	3	

2025-09-09 10:00:00

1202

	Init	36683	36792	+	0
	Intr	36895	36996	+	0
	Term	37400	38513	+	0
5	>3510339	/4190			
	len =	570	nex =	1	
10	Sngl	37975	38544	+	0
	>3510339	/27885			
	len =	910	nex =	4	
15	Init	47055	47312	+	0
	Intr	47405	47463	+	0
	Intr	47558	47675	+	0
	Term	47755	47961	+	0
20	>3510339	/32353			
	len =	1336	nex =	5	
25	Init	53149	53304	+	0
	Intr	53397	53504	+	0
	Intr	53574	53700	+	0
	Intr	53970	54038	+	0
	Term	54133	54484	+	0
30	>3510339	/14885			
	len =	688	nex =	1	
35	Sngl	8005	7318	-	0
	>3510340	/126460			
	len =	934	nex =	2	
40	Term	21669	21324	-	0
	Init	22257	21750	-	0
	>3510340	/40096			
	len =	2290	nex =	7	
	Term	38031	37656	-	0
	Intr	38245	38120	-	0
50	Intr	38564	38358	-	0
	Intr	38726	38658	-	0
	Intr	38905	38813	-	0
	Intr	39170	38997	-	0
	Init	39945	39536	-	0
55	>3510340	/14006			
	len =	3759	nex =	18	
60	Term	41382	41098	-	0

1203

	Intr	41552	41486	-	0
	Intr	41808	41760	-	0
	Intr	42179	42123	-	0
5	Intr	42406	42335	-	0
	Intr	42585	42494	-	0
	Intr	42747	42675	-	0
	Intr	42944	42870	-	0
	Intr	43109	43045	-	0
	Intr	43297	43210	-	0
10	Intr	43431	43387	-	0
	Intr	43733	43614	-	0
	Intr	43878	43840	-	0
	Intr	44054	43971	-	0
	Intr	44229	44138	-	0
15	Intr	44386	44312	-	0
	Intr	44577	44484	-	0
	Init	44856	44655	-	0
	>3510341 /10879				
20	len =	1930	nex =	0	
	>3510341 /36005				
25	len =	691	nex =	1	
	Sngl	40942	41632	+	0
	>3510341 /32092				
30	len =	2099	nex =	9	
	Term	48160	47836	-	0
	Intr	48310	48244	-	0
35	Intr	48500	48383	-	0
	Intr	48656	48600	-	0
	Intr	48836	48741	-	0
	Intr	49030	48923	-	0
	Intr	49183	49114	-	0
40	Intr	49608	49481	-	0
	Init	49934	49746	-	0
	>3510341 /18565				
45	len =	1332	nex =	3	
	Term	55802	55320	-	0
	Intr	56319	56196	-	0
	Init	56651	56408	-	0
50	>3510342 /21714				
	len =	1319	nex =	1	
55	Sngl	54886	56204	+	0
	>3510342 /24640				
60	len =	730	nex =	3	

1204

	Init	56678	56815	+	0
	Intr	56912	56980	+	0
	Term	57073	57398	+	0
5	>3510342	/123863			
	len =	910	nex =	1	
	Sngl	60334	61240	+	0
10	>3510342	/534			
	len =	536	nex =	1	
15	Sngl	9075	8540	-	0
	>3510343	/40742			
	len =	1891	nex =	7	
20	Init	13415	13697	+	0
	Intr	13866	14012	+	0
	Intr	14098	14196	+	0
	Intr	14285	14476	+	0
25	Intr	14612	14731	+	0
	Intr	14818	14889	+	0
	Term	14970	15305	+	0
	>3510343	/24171			
30	len =	1330	nex =	1	
	Sngl	2256	3585	+	0
35	>3510343	/2722			
	len =	503	nex =	1	
	Sngl	30214	29712	-	0
40	>3510343	/106836			
	len =	1702	nex =	9	
45	Term	33894	33628	-	0
	Intr	34089	33998	-	0
	Intr	34258	34188	-	0
	Intr	34371	34329	-	0
	Intr	34518	34462	-	0
50	Intr	34704	34605	-	0
	Intr	34927	34777	-	0
	Intr	35103	35021	-	0
	Init	35329	35222	-	0
55	>3510343	/6903			
	len =	1763	nex =	4	
	Init	37722	38574	+	0
60	Intr	38655	38839	+	0

1205

	Intr	38929	39039	+	0
	Term	39124	39484	+	0
	>3510343 /347				
5	len =	1651	nex =	3	
	Init	4180	4489	+	0
	Intr	5552	5687	+	0
10	Term	5778	5830	+	0
	>3510343 /97314				
	len =	1179	nex =	3	
15	Term	41393	41220	-	0
	Intr	41794	41558	-	0
	Init	42398	42147	-	0
	>3510343 /29713				
20	len =	614	nex =	1	
	Sngl	43022	42748	-	0
25	>3510343 /42932				
	len =	3299	nex =	12	
30	Term	44028	43691	-	0
	Intr	44238	44168	-	0
	Intr	44426	44307	-	0
	Intr	44873	44679	-	0
	Intr	45135	45071	-	0
35	Intr	45321	45231	-	0
	Intr	45556	45495	-	0
	Intr	45732	45682	-	0
	Intr	45918	45817	-	0
	Intr	46113	46006	-	0
40	Intr	46273	46210	-	0
	Init	46989	46641	-	0
	>3510343 /27758				
45	len =	1726	nex =	1	
	Sngl	48931	50656	+	0
	>3510343 /41873				
50	len =	1120	nex =	1	
	Sngl	49483	50602	+	0
	>3510343 /103302				
55	len =	346	nex =	1	
	Sngl	50315	50660	+	0
60					

2025 RELEASE UNDER E.O. 14176

1206

>3510343 /11988

len = 2000 nex = 1

5 Sngl 53425 52376 - 0

>3510343 /40911

len = 533 nex = 2

10 Term 54279 53987 - 0
Init 54519 54367 - 0

>3510343 /4645

15 len = 2130 nex = 6

20 Init 68383 68789 + 0
Intr 69020 69154 + 0
Intr 69249 69327 + 0
Intr 69422 69516 + 0
Intr 69735 69902 + 0
Term 69982 70512 + 0

25 >3510343 /40013

len = 219 nex = 1

30 Sngl 77830 77612 - 0

>3510343 /4267

len = 790 nex = 1

35 Sngl 82643 83428 + 0

>3510344 /2679

len = 511 nex = 1

40 Sngl 43185 43695 + 0

>3510345 /38935

45 len = 1870 nex = 5

50 Term 20537 19806 - 0
Intr 20684 20623 - 0
Intr 20927 20816 - 0
Intr 21079 21017 - 0
Init 21667 21540 - 0

>3510345 /16321

55 len = 1839 nex = 6

60 Init 30949 31052 + 0
Intr 31463 31499 + 0
Intr 31947 32020 + 0
Intr 32124 32217 + 0

2025 RELEASE UNDER E.O. 14176

1207

Intr	32330	32412	+	0
Term	32494	32787	+	0

>3510345 /9887

5	len =	1957	nex =	5
---	-------	------	-------	---

	Init	31463	31499	+	0
	Intr	31947	32020	+	0
10	Intr	32124	32217	+	0
	Intr	32330	32412	+	0
	Term	32494	32923	+	0

>3510345 /40590

15	len =	2650	nex =	7
----	-------	------	-------	---

	Init	36093	36290	+	0
	Intr	36588	37160	+	0
20	Intr	37250	37316	+	0
	Intr	37733	37824	+	0
	Intr	37917	38126	+	0
	Intr	38223	38360	+	0
	Term	38470	38726	+	0

>3510345 /42141

	len =	3010	nex =	12
--	-------	------	-------	----

30	Init	61602	61918	+	0
	Intr	62116	62244	+	0
	Intr	62321	62413	+	0
	Intr	62503	62582	+	0
	Intr	62663	62776	+	0
35	Intr	62944	63017	+	0
	Intr	63156	63232	+	0
	Intr	63328	63411	+	0
	Intr	63507	63599	+	0
	Intr	63687	63794	+	0
40	Intr	63877	63948	+	0
	Term	64039	64603	+	0

>3510345 /21808

45	len =	3836	nex =	4
----	-------	------	-------	---

	Term	82161	81264	-	0
	Intr	82858	82787	-	0
	Intr	83396	83347	-	0
50	Init	84608	84277	-	0

>3510346 /34019

	len =	465	nex =	1
--	-------	-----	-------	---

55	Sngl	19046	18582	-	0
----	------	-------	-------	---	---

>3510346 /3859

60	len =	583	nex =	1
----	-------	-----	-------	---

2025 RELEASE UNDER E.O. 14176

1208

	Sngl	19162	18580	-	0
	>3510346	/41105			
5	len =	2655	nex =	4	
	Term	21838	21322	-	0
	Intr	22141	21929	-	0
10	Intr	22425	22322	-	0
	Init	23176	22513	-	0
	>3510346	/18886			
15	len =	1400	nex =	1	
	Sngl	2565	1166	-	0
	>3510346	/6830			
20	len =	1411	nex =	2	
	Init	28915	29605	+	0
	Term	29936	30325	+	0
25	>3510346	/19036			
	len =	925	nex =	2	
30	Init	29413	29605	+	0
	Term	29936	30323	+	0
	>3510347	/32287			
35	len =	1572	nex =	6	
	Term	17536	17338	-	0
	Intr	17706	17628	-	0
	Intr	17947	17881	-	0
40	Intr	18159	18060	-	0
	Intr	18386	18246	-	0
	Init	18909	18664	-	0
	>3510347	/10832			
45	len =	1590	nex =	5	
	Term	17536	17344	-	0
	Intr	17706	17628	-	0
50	Intr	17947	17881	-	0
	Intr	18159	18060	-	0
	Init	18386	18246	-	0
	>3510347	/8265			
55	len =	1462	nex =	1	
	Sngl	25560	25814	+	0
60	>3510347	/37190			

2025 RELEASE UNDER E.O. 14176

1209

	len =	2683	nex =	8	
	Init	26268	26469	+	0
5	Intr	27118	27309	+	0
	Intr	27478	27554	+	0
	Intr	27755	27814	+	0
	Intr	27939	28142	+	0
	Intr	28258	28320	+	0
10	Intr	28413	28481	+	0
	Term	28719	28950	+	0
>3510347		/21563			
15	len =	1497	nex =	6	
	Term	32966	32764	-	0
	Intr	33122	33065	-	0
	Intr	33434	33389	-	0
20	Intr	33570	33535	-	0
	Intr	33921	33809	-	0
	Init	34260	34059	-	0
>3510347		/6			
25	len =	2248	nex =	8	
	Term	35112	34642	-	0
	Intr	35312	35190	-	0
30	Intr	35520	35412	-	0
	Intr	35687	35606	-	0
	Intr	36105	35898	-	0
	Intr	36365	36207	-	0
	Intr	36591	36457	-	0
35	Init	36889	36687	-	0
>3510347		/40282			
40	len =	1030	nex =	1	
	Sngl	44145	43121	-	0
>3510347		/21002			
45	len =	1112	nex =	4	
	Init	44576	44637	+	0
	Intr	44840	44941	+	0
	Intr	45017	45138	+	0
50	Term	45230	45687	+	0
>3510347		/8259			
55	len =	911	nex =	2	
	Init	46905	47248	+	0
	Term	47327	47815	+	0
60	>3510347		/101505		

1210

	len =	1259	nex =	5	
	Init	49449	49599	+	0
	Intr	49835	50106	+	0
5	Intr	50195	50285	+	0
	Intr	50364	50430	+	0
	Term	50515	50707	+	0
	>3510347 /2528				
10	len =	1377	nex =	2	
	Term	57801	57178	-	0
	Init	58554	58264	-	0
15	>3510347 /112017				
	len =	914	nex =	2	
20	Term	57801	57641	-	0
	Init	58554	58264	-	0
	>3510347 /21882				
25	len =	2650	nex =	3	
	Init	60423	61009	+	0
	Intr	61588	61725	+	0
	Term	62599	63070	+	0
30	>3510347 /93510				
	len =	1477	nex =	3	
35	Init	64300	64751	+	0
	Intr	64887	65024	+	0
	Term	65353	65776	+	0
	>3510347 /25205				
40	len =	2153	nex =	5	
	Init	68370	68963	+	0
	Intr	69054	69194	+	0
45	Intr	69282	69422	+	0
	Intr	69950	70056	+	0
	Term	70149	70522	+	0
	>3510347 /10				
50	len =	731	nex =	2	
	Term	6348	6208	-	0
	Init	6938	6424	-	0
55	>3510347 /20808				
	len =	1651	nex =	4	
60	Init	72240	72762	+	0

1211

Intr	72870	73010	+	0
Intr	73158	73298	+	0
Term	73581	73890	+	0

5 >3510347 /3268

len = 670 nex = 2

Term	74338	73874	-	0
Init	74537	74427	-	0

>3510347 /34224

15 len = 2512 nex = 6

Init	8462	8802	+	0
Intr	9019	9104	+	0
Intr	9299	9349	+	0
Intr	10221	10287	+	0
Intr	10414	10493	+	0
Term	10655	10973	+	0

>3513725 /36053

25 len = 1750 nex = 7

Init	13019	13159	+	0
Intr	13469	13609	+	0
Intr	13779	13816	+	0
Intr	13897	13947	+	0
Intr	14038	14205	+	0
Intr	14282	14394	+	0
Term	14473	14764	+	0

35 >3513725 /93362

len = 2622 nex = 6

Init	25174	25560	+	0
Intr	25648	25712	+	0
Intr	25798	25884	+	0
Intr	25990	26129	+	0
Intr	27220	27274	+	0
Term	27422	27795	+	0

>3513725 /30013

len = 2650 nex = 6

Init	25215	25560	+	0
Intr	25648	25712	+	0
Intr	25798	25884	+	0
Intr	25990	26129	+	0
Intr	27220	27274	+	0
Term	27422	27859	+	0

>3513725 /15608

len = 2626 nex = 9

60

1212

	Init	28826	29089	+	0
	Intr	29751	29800	+	0
	Intr	29928	30063	+	0
5	Intr	30157	30258	+	0
	Intr	30340	30567	+	0
	Intr	30642	30720	+	0
	Intr	30807	30889	+	0
	Intr	30975	31066	+	0
10	Term	31194	31451	+	0
	>3513725	/11425			
	len =	1812	nex =	3	
15	Init	43000	43341	+	0
	Intr	43510	43706	+	0
	Term	44140	44588	+	0
	>3513725	/4588			
20	len =	1847	nex =	3	
	Init	42785	43341	+	0
	Intr	43510	43706	+	0
25	Term	44140	44631	+	0
	>3513725	/19449			
	len =	1469	nex =	3	
30	Init	48183	48572	+	0
	Intr	48751	48856	+	0
	Term	49391	49651	+	0
35	>3513725	/33161			
	len =	1990	nex =	5	
40	Term	70844	70396	-	0
	Intr	71145	71026	-	0
	Intr	71729	71348	-	0
	Intr	72167	72131	-	0
	Init	72376	72258	-	0
45	>3513725	/2538			
	len =	2710	nex =	4	
50	Term	88530	88065	-	0
	Intr	88919	88628	-	0
	Intr	89919	89306	-	0
	Init	90771	90027	-	0
55	>3522932	/1233			
	len =	468	nex =	1	
	Sngl	14882	14415	-	0
60	>3522932	/6247			

1213

5	len =	629	nex =	1	
	Sngl	14910	14282	-	0
	>3522932	/34004			
10	len =	670	nex =	1	
	Sngl	14939	14277	-	0
	>3522932	/34114			
15	len =	970	nex =	1	
	Sngl	26932	27893	+	0
	>3522932	/119203			
20	len =	310	nex =	1	
	Sngl	31113	31414	+	0
	>3522932	/40043			
25	len =	2092	nex =	2	
	Init	33181	34387	+	0
	Term	34475	35272	+	0
30	>3522932	/17422			
	len =	1694	nex =	8	
	Term	85840	85605	-	0
35	Intr	85984	85923	-	0
	Intr	86150	86064	-	0
	Intr	86316	86236	-	0
40	Intr	86490	86395	-	0
	Intr	86710	86576	-	0
	Intr	86925	86803	-	0
45	Init	87172	87000	-	0
	>3548797	/39349			
	len =	1390	nex =	1	
50	Sngl	45745	47127	+	0
	>3548797	/118505			
	len =	1236	nex =	5	
55	Term	54662	54394	-	0
	Intr	54808	54746	-	0
	Intr	54939	54898	-	0
60	Intr	55250	55035	-	0
	Init	55629	55345	-	0
	>3548797	/37655			

1214

	len =	2848	nex =	11	
	Term	73586	73293	-	0
5	Intr	73848	73771	-	0
	Intr	74084	73948	-	0
	Intr	74268	74199	-	0
	Intr	74454	74368	-	0
	Intr	74621	74571	-	0
10	Intr	74888	74761	-	0
	Intr	75233	75148	-	0
	Intr	75468	75317	-	0
	Intr	75678	75563	-	0
	Init	76140	75987	-	0
15					
	>3582315	/29796			
	len =	1556	nex =	1	
20	Sngl	45020	43465	-	0
	>3582315	/34214			
	len =	1941	nex =	6	
25	Init	46259	46498	+	0
	Intr	46588	46734	+	0
	Intr	46826	46878	+	0
	Intr	46958	47120	+	0
30	Intr	47207	47435	+	0
	Term	47518	48199	+	0
	>3582315	/12086			
35	len =	651	nex =	1	
	Sngl	51096	50446	-	0
	>3582315	/36220			
40	len =	1908	nex =	2	
	Init	64549	65215	+	0
	Term	66042	66456	+	0
45					
	>3600029	/26159			
	len =	390	nex =	1	
50	Sngl	24328	23939	-	0
	>3600029	/35159			
	len =	1179	nex =	1	
55	Sngl	25177	23999	-	0
	>3600045	/24130			
60	len =	1030	nex =	3	

1215

	Term	15293	14783	-	0
	Intr	15574	15428	-	0
	Init	15807	15656	-	0
5	>3600045 /32602				
	len =	1450	nex =	1	
10	Sngl	25383	26825	+	0
	>3600045 /25988				
	len =	824	nex =	2	
15	Term	3124	2675	-	0
	Init	3498	3344	-	0
	>3600045 /22850				
20	len =	408	nex =	2	
	Term	41689	41456	-	0
	Init	41863	41773	-	0
25	>3600045 /35662				
	len =	1570	nex =	3	
30	Term	51918	51401	-	0
	Intr	52566	51996	-	0
	Init	52969	52885	-	0
	>3608126 /5901				
35	len =	1296	nex =	3	
	Term	11196	10881	-	0
	Intr	11486	11310	-	0
40	Init	12176	11888	-	0
	>3608126 /3992				
	len =	1349	nex =	5	
45	Term	26268	26001	-	0
	Intr	26428	26385	-	0
	Intr	26650	26563	-	0
	Intr	27078	27031	-	0
50	Init	27343	27172	-	0
	>3608126 /9670				
	len =	1059	nex =	2	
55	Term	1939	1719	-	0
	Init	2777	2477	-	0
	>3608126 /17424				
60					

1216

	len =	2011	nex =	8	
	Term	33042	32798	-	0
	Intr	33307	33131	-	0
5	Intr	33518	33400	-	0
	Intr	33658	33592	-	0
	Intr	33803	33742	-	0
	Intr	34105	33911	-	0
	Intr	34389	34199	-	0
10	Init	34808	34562	-	0
	>3608126	/38584			
	len =	1669	nex =	3	
15	Init	34965	35667	+	0
	Intr	35839	36034	+	0
	Term	36222	36633	+	0
20	>3608126	/13821			
	len =	873	nex =	3	
	Init	39040	39422	+	0
25	Intr	39532	39576	+	0
	Term	39671	39912	+	0
	>3608126	/36332			
30	len =	3702	nex =	6	
	Init	4194	4297	+	0
	Intr	5595	5680	+	0
	Intr	6594	6973	+	0
35	Intr	7084	7213	+	0
	Intr	7302	7588	+	0
	Term	7672	7895	+	0
	>3608126	/23118			
40	len =	814	nex =	4	
	Term	56920	56791	-	0
	Intr	57180	57010	-	0
45	Intr	57415	57327	-	0
	Init	57604	57495	-	0
	>3608126	/24194			
50	len =	1115	nex =	4	
	Term	56920	56626	-	0
	Intr	57180	57010	-	0
	Intr	57415	57327	-	0
55	Init	57740	57495	-	0
	>3617740	/23672			
60	len =	2710	nex =	6	

1217

	Term	779	523	-	0
	Intr	995	869	-	0
	Intr	1176	1079	-	0
	Intr	1629	1462	-	0
5	Intr	2092	1983	-	0
	Init	2233	2181	-	0

>3617740 /14534

10	len =	2050	nex =	5	
	Init	38982	39082	+	0
	Intr	39269	39379	+	0
	Intr	39472	39776	+	0
15	Intr	39865	40274	+	0
	Term	40379	40796	+	0

>3617740 /17803

20	len =	878	nex =	4	
	Init	38750	38796	+	0
	Intr	38982	39082	+	0
	Intr	39269	39379	+	0
25	Term	39472	39614	+	0

>3617740 /125212

30	len =	613	nex =	2	
	Init	40232	40274	+	0
	Term	40379	40844	+	0

>3617740 /4609

35	len =	3370	nex =	14	
	Term	47827	47264	-	0
	Intr	48084	47941	-	0
40	Intr	48277	48170	-	0
	Intr	48449	48436	-	0
	Intr	48556	48515	-	0
	Intr	48691	48640	-	0
45	Intr	48810	48769	-	0
	Intr	48971	48922	-	0
	Intr	49144	49080	-	0
	Intr	49457	49334	-	0
	Intr	49729	49648	-	0
50	Intr	49877	49801	-	0
	Intr	50044	49973	-	0
	Init	50199	50135	-	0

>3641835 /124015

55	len =	586	nex =	2	
	Term	104751	104680	-	0
	Init	105265	105103	-	0

60 >3641835 /24998

1218

	len =	1358	nex =	3	
5	Term	104368	103951	-	0
	Intr	104751	104466	-	0
	Init	105308	105103	-	0
	>3641835	/40208			
10	len =	1313	nex =	3	
	Term	106166	106021	-	0
	Intr	106634	106461	-	0
	Init	107333	106949	-	0
15	>3641835	/10510			
	len =	692	nex =	1	
20	Sngl	118544	117853	-	0
	>3641835	/29665			
25	len =	2207	nex =	6	
	Term	13437	13279	-	0
	Intr	13643	13530	-	0
	Intr	14090	13733	-	0
	Intr	14763	14339	-	0
30	Intr	15097	15003	-	0
	Init	15485	15170	-	0
	>3641835	/101430			
35	len =	600	nex =	3	
	Term	19477	19405	-	0
	Intr	19845	19550	-	0
	Init	20004	19937	-	0
40	>3641835	/8586			
	len =	1191	nex =	3	
45	Term	40504	40292	-	0
	Intr	41034	40900	-	0
	Init	41253	41179	-	0
	>3641835	/40231			
50	len =	1140	nex =	1	
	Sngl	57915	59054	+	0
55	>3641835	/6475			
	len =	460	nex =	1	
60	Sngl	58753	59203	+	0

1219

>3641835 /42589

len = 1193 nex = 3

5	Term	71259	70935	-	0
	Intr	71875	71830	-	0
	Init	72127	71973	-	0

>3641835 /29919

10 len = 1271 nex = 3

	Term	71259	70897	-	0
	Intr	71875	71830	-	0
15	Init	72167	71973	-	0

>3641835 /27483

20 len = 833 nex = 1

Sngl 73741 72909 - 0

>3641835 /20883

25 len = 797 nex = 1

Sngl 73759 72963 - 0

>3641835 /28635

30 len = 3595 nex = 12

	Init	74327	74696	+	0
	Intr	75055	75097	+	0
35	Intr	75428	75503	+	0
	Intr	75665	75734	+	0
	Intr	76009	76158	+	0
	Intr	76267	76371	+	0
40	Intr	76474	76620	+	0
	Intr	76796	76871	+	0
	Intr	77032	77120	+	0
	Intr	77218	77309	+	0
	Intr	77476	77518	+	0
45	Term	77612	77921	+	0

>3641835 /37733

len = 2501 nex = 13

50	Init	78502	78584	+	0
	Intr	78669	78711	+	0
	Intr	78845	78934	+	0
	Intr	79022	79097	+	0
55	Intr	79225	79294	+	0
	Intr	79399	79548	+	0
	Intr	79650	79754	+	0
	Intr	79835	79981	+	0
	Intr	80095	80170	+	0
60	Intr	80300	80388	+	0
	Intr	80465	80556	+	0

202502261500

1220

Intr	80724	80766	+	0
Term	80863	80994	+	0

Term	80863	80994	+	0
------	-------	-------	---	---

>3641835 /32357

```

5      len =      3190      nex =      10

```

Term	81717	81263	-	0
------	-------	-------	---	---

Intr	81922	81848	-	0
------	-------	-------	---	---

10	Intr	82305	82153	-	0
----	------	-------	-------	---	---

Intr	82488	82426	-	0
------	-------	-------	---	---

Intr	82779	82579	-	0
------	-------	-------	---	---

Intr	82915	82915	-	0
------	-------	-------	---	---

Intr	83421	83270	-	0
------	-------	-------	---	---

15	Intr	83600	83531	-	0
----	------	-------	-------	---	---

Intr	83792	83679	-	0
------	-------	-------	---	---

Init	84448	84248	-	0
------	-------	-------	---	---

>3641835 /143247

```
20      len =      1363      nex =          5
```

Init	85011	85116	+	0
------	-------	-------	---	---

Intr	85354	85491	+	0
------	-------	-------	---	---

25	Intr	85562	85747	+	0
----	------	-------	-------	---	---

Intr	85840	86040	+	0
------	-------	-------	---	---

Term	86137	86234	+	0
------	-------	-------	---	---

>3641835 /24125

```

30      len =      578      nex =      2

```

Init	85840	86040	+	0
------	-------	-------	---	---

Term	86137	86267	+	0
------	-------	-------	---	---

35
 >3641835 /27838

```
len = 1964    nex = 4
```

40 Tnit 92854 93018 + 0

Intr	93122	93163	+	(
------	-------	-------	---	---

Intr	93954	94035	+	0
------	-------	-------	---	---

Term	94181	94236	+	0
------	-------	-------	---	---

45 >3641835 /18715

```
len = 2001    nex = 4
```

Init	92859	93018	+	0
------	-------	-------	---	---

50	Intr	93122	93163	+	
----	------	-------	-------	---	--

Intr	93954	94035	+	
------	-------	-------	---	--

Term	94181	94236	+	
------	-------	-------	---	--

>3641835 /22619

```
55      len =      837      nex =      1
```

Snq1	97080	96244	-
------	-------	-------	---

60 >3643588 /7152

1221

	len =	700	nex =	1	
5	Sngl	100213	100912	+	0
	>3643588	/26655			
	len =	2932	nex =	10	
10	Init	23965	24029	+	0
	Intr	24558	24733	+	0
	Intr	24818	24949	+	0
	Intr	25037	25087	+	0
	Intr	25184	25280	+	0
15	Intr	25536	25582	+	0
	Intr	25690	25909	+	0
	Intr	26128	26257	+	0
	Intr	26350	26392	+	0
	Term	26519	26896	+	0
20	>3643588	/39893			
	len =	1292	nex =	1	
25	Sngl	34172	32881	-	0
	>3643588	/33811			
	len =	619	nex =	1	
30	Sngl	38584	37966	-	0
	>3643588	/14229			
35	len =	88	nex =	1	
	Sngl	44849	44762	-	0
	>3643588	/107487			
40	len =	688	nex =	1	
	Sngl	91073	91760	+	0
45	>3650026	/105053			
	len =	1096	nex =	1	
	Sngl	15351	15025	-	0
50	>3650026	/41621			
	len =	1450	nex =	2	
55	Term	14294	13985	-	0
	Init	15430	15025	-	0
	>3650026	/98613			
60	len =	103	nex =	1	

1222

	Sngl	20145	20247	+	0
	>3650026	/1845			
5	len =	1078	nex =	1	
	Sngl	25288	25744	+	0
10	>3650026	/105106			
	len =	430	nex =	1	
	Sngl	25301	25714	+	0
15	>3650026	/18430			
	len =	2276	nex =	5	
20	Term	27554	27148	-	0
	Intr	27977	27733	-	0
	Intr	28182	28075	-	0
	Intr	28573	28309	-	0
	Init	29423	29071	-	0
25	>3659491	/25383			
	len =	4510	nex =	13	
30	Term	9461	9027	-	0
	Intr	9645	9554	-	0
	Intr	9862	9749	-	0
	Intr	10492	10129	-	0
	Intr	10829	10696	-	0
35	Intr	11104	10996	-	0
	Intr	11365	11244	-	0
	Intr	11551	11442	-	0
	Intr	11745	11641	-	0
	Intr	12001	11843	-	0
40	Intr	12392	12337	-	0
	Intr	12653	12602	-	0
	Init	13533	13069	-	0
	>3659491	/40893			
45	len =	1515	nex =	4	
	Init	17945	18140	+	0
	Intr	18495	18596	+	0
50	Intr	18711	18833	+	0
	Term	19154	19459	+	0
	>3659491	/11713			
55	len =	647	nex =	1	
	Sngl	29537	28891	-	0
	>3659491	/1822			
60					

1223

	len =	1542	nex =	4	
	Term	30282	29975	-	0
	Intr	30669	30547	-	0
5	Intr	30870	30769	-	0
	Init	31516	31375	-	0
	>3659491 /16879				
10	len =	792	nex =	2	
	Term	43340	43008	-	0
	Init	43799	43512	-	0
15	>3659491 /31696				
	len =	1488	nex =	4	
	Init	52934	53129	+	0
20	Intr	53228	53290	+	0
	Intr	54065	54124	+	0
	Term	54224	54421	+	0
	>3659491 /3853				
25	len =	1674	nex =	3	
	Init	67957	68461	+	0
	Intr	68545	68737	+	0
30	Term	69218	69630	+	0
	>3659491 /22630				
	len =	1510	nex =	3	
35	Init	67989	68461	+	0
	Intr	68545	68737	+	0
	Term	69218	69495	+	0
40	>3659491 /119226				
	len =	1460	nex =	1	
	Sngl	80172	80692	+	0
45	>3659491 /36272				
	len =	1398	nex =	1	
50	Sngl	80172	80692	+	0
	>3659491 /31768				
	len =	1795	nex =	3	
55	Init	82901	82956	+	0
	Intr	83626	83993	+	0
	Term	84259	84695	+	0
60	>3659491 /1530				

1224

	len =	1113	nex =	2	
	Term	86548	85965	-	0
5	Init	87077	86859	-	0
	>3659491 /17503				
	len =	270	nex =	1	
10	Sngl	9313	9044	-	0
	>3668073 /16622				
15	len =	1358	nex =	4	
	Init	23250	23680	+	0
	Intr	23767	23864	+	0
	Intr	23960	24025	+	0
20	Term	24310	24607	+	0
	>3668073 /37897				
	len =	1111	nex =	1	
25	Sngl	51343	52453	+	0
	>3668073 /37026				
30	len =	1930	nex =	6	
	Init	694	789	+	0
	Intr	1152	1397	+	0
	Intr	1494	1661	+	0
35	Intr	1741	1834	+	0
	Intr	1915	2085	+	0
	Term	2221	2618	+	0
	>3687221 /16515				
40	len =	1589	nex =	3	
	Init	102216	102407	+	0
	Intr	102664	102801	+	0
45	Term	103376	103791	+	0
	>3687221 /9466				
	len =	1096	nex =	3	
50	Term	107635	107345	-	0
	Intr	108241	108104	-	0
	Init	108440	108355	-	0
55	>3687221 /14718				
	len =	1498	nex =	1	
	Sngl	121547	121349	-	0
60					

1225

>3687221 /15861

len = 973 nex = 5

5	Init	91525	91626	+	0
	Intr	91730	91799	+	0
	Intr	91913	91991	+	0
	Intr	92096	92177	+	0
	Term	92263	92497	+	0

10

>3687221 /5064

len = 1290 nex = 4

15	Term	95134	94875	-	0
	Intr	95305	95237	-	0
	Intr	95573	95487	-	0
	Init	96164	95953	-	0

20

>3688169 /36835

len = 2308 nex = 8

25	Term	10054	9794	-	0
	Intr	10172	10133	-	0
	Intr	10479	10407	-	0
	Intr	10683	10607	-	0
	Intr	10809	10773	-	0
	Intr	10942	10882	-	0
30	Intr	11425	11225	-	0
	Init	12081	11998	-	0

>3688169 /39326

35 len = 1424 nex = 4

	Init	16580	16851	+	0
	Intr	16934	16974	+	0
	Intr	17077	17132	+	0
40	Term	17802	18003	+	0

>3688169 /8739

45 len = 1450 nex = 4

	Init	29897	30039	+	0
	Intr	30457	30564	+	0
	Intr	30834	31001	+	0
	Term	31120	31345	+	0

50

>3688169 /95066

len = 850 nex = 1

55 Sngl 6958 7799 + 0

>3688169 /4868

len = 590 nex = 1

60

20250916 09:25:03

090806

					1226	0
	Sngl	72372	71783	-		
	>3695372	/25253				
5	len =	2094	nex =	11		
	Init	1783	2087	+		0
	Intr	2177	2257	+		0
10	Intr	2345	2440	+		0
	Intr	2533	2616	+		0
	Intr	2705	2774	+		0
	Intr	2854	2936	+		0
	Intr	3017	3096	+		0
15	Intr	3178	3276	+		0
	Intr	3365	3434	+		0
	Term	3708	3876	+		0
	>3695372	/20974				
20	len =	9399	nex =	1		
	Sngl	20229	19834	-		0
25	>3695386	/24263				
	len =	711	nex =	1		
	Sngl	19462	20172	+		0
30	>3695400	/34605				
	len =	462	nex =	1		
35	Sngl	10570	11031	+		0
	>3695400	/1489				
40	len =	1587	nex =	2		
	Term	15637	15331	-		0
	Init	16917	16616	-		0
	>3695400	/10189				
45	len =	985	nex =	2		
	Init	25271	25588	+		0
	Term	25943	26252	+		0
50	>3695400	/1831				
	len =	477	nex =	1		
55	Sngl	26645	26169	-		0
	>3695400	/35152				
60	len =	455	nex =	1		

1227

Snq1	27846	28300	+	0
------	-------	-------	---	---

>3695400 /108385

```
5      len =      629      nex =      3
```

Init	3404	3506	+	0
------	------	------	---	---

Intr	3589	3680	+	0
------	------	------	---	---

Term	3749	4032	+	0
------	------	------	---	---

10

>3695400 /2125

```
len = 1522    nex = 2
```

15	Term	5925	5502	-	0
----	------	------	------	---	---

Init	7023	6790	-	0
------	------	------	---	---

>3695400 /28203

```
20      len =      2144      nex =          5
```

Term	77122	76779	-	0
------	-------	-------	---	---

Intr	77397	77265	-	0
------	-------	-------	---	---

Intr	77620	77515	-	0
------	-------	-------	---	---

25	Intr	77799	77729	-	0
----	------	-------	-------	---	---

Init	78922	78765	-	0
------	-------	-------	---	---

>3702315 /32891

```
30      len =      2096      nex =          9
```

Init	15705	15815	+	0
------	-------	-------	---	---

Intr	15919	16138	+	0
------	-------	-------	---	---

Intr	16274	16489	+	0
------	-------	-------	---	---

35	Intr	16586	16719	+	0
----	------	-------	-------	---	---

Intr	16845	16916	+	0
------	-------	-------	---	---

Intr	17042	17131	+	0
------	-------	-------	---	---

Intr	17227	17313	+	(
------	-------	-------	---	---

Intr	17401	17480	+	0
------	-------	-------	---	---

40 Term 17583 17792 + 0

>3702315 /34383

```
len = 2082 nex = 7
```

Init	18147	18358	+
------	-------	-------	---

Intr	18507	18722	+
------	-------	-------	---

Intr	18818	18951	+
------	-------	-------	---

Intr	19279	19368	+
------	-------	-------	---

50	Intr	19462	19548	+
----	------	-------	-------	---

Intr	19649	19728	+
------	-------	-------	---

Term	19837	20078	+
------	-------	-------	---

>3702315 /38881

```
55      len =      2006      nex =      7
```

Init	18051	18358	+
------	-------	-------	---

Intr	18507	18722	+
------	-------	-------	---

60	Intr	18818	18951	+
----	------	-------	-------	---

1228

	Intr	19279	19368	+	0
	Intr	19462	19548	+	0
	Intr	19649	19728	+	0
	Term	19837	20056	+	0
5	>3702315 /126587				
	len =	1758	nex =	9	
10	Init	37610	37788	+	0
	Intr	37865	37924	+	0
	Intr	38002	38107	+	0
	Intr	38260	38302	+	0
	Intr	38382	38454	+	0
15	Intr	38543	38630	+	0
	Intr	38770	38826	+	0
	Intr	38917	38995	+	0
	Term	39096	39367	+	0
20	>3702315 /39624				
	len =	1673	nex =	9	
	Init	37636	37788	+	0
25	Intr	37865	37924	+	0
	Intr	38002	38107	+	0
	Intr	38260	38302	+	0
	Intr	38382	38454	+	0
	Intr	38543	38630	+	0
30	Intr	38770	38826	+	0
	Intr	38917	38995	+	0
	Term	39096	39308	+	0
35	>3702315 /1835				
	len =	1311	nex =	8	
	Init	37685	37788	+	0
	Intr	37865	37924	+	0
40	Intr	38002	38107	+	0
	Intr	38260	38302	+	0
	Intr	38382	38454	+	0
	Intr	38543	38630	+	0
	Intr	38770	38826	+	0
45	Term	38917	38989	+	0
	>3702315 /13319				
	len =	2328	nex =	8	
50	Term	39649	39295	-	0
	Intr	39862	39780	-	0
	Intr	40012	39960	-	0
	Intr	40165	40081	-	0
	Intr	40534	40435	-	0
55	Intr	40776	40626	-	0
	Intr	40939	40857	-	0
	Init	41622	41363	-	0
60	>3702315 /16400				

1229

	len =	746	nex =	2	
	Init	4313	4574	+	0
	Term	4732	5058	+	0
5	>3702315 /22834				
	len =	711	nex =	2	
10	Init	4314	4711	+	0
	Term	4732	5024	+	0
	>3702315 /2011				
15	len =	815	nex =	1	
	Sngl	50261	49447	-	0
	>3702315 /150696				
20	len =	813	nex =	1	
	Sngl	805	1395	+	0
25	>3702315 /9694				
	len =	1390	nex =	2	
30	Term	63455	63085	-	0
	Init	64470	64155	-	0
	>3702315 /2747				
35	len =	1450	nex =	2	
	Term	63455	63044	-	0
	Init	64488	64155	-	0
40	>3702315 /37367				
	len =	1800	nex =	6	
	Init	75775	75964	+	0
	Intr	76051	76180	+	0
45	Intr	76334	76471	+	0
	Intr	76560	76892	+	0
	Intr	76989	77172	+	0
	Term	77268	77574	+	0
50	>3702722 /2662				
	len =	2085	nex =	9	
	Term	14429	14277	-	0
55	Intr	14686	14527	-	0
	Intr	14858	14767	-	0
	Intr	15138	14935	-	0
	Intr	15415	15221	-	0
	Intr	15543	15495	-	0
60	Intr	15749	15646	-	0

1230

Intr	15877	15838	-	0
Init	16361	16286	-	0

>3702722 /34554

5	len =	2372	nex =	9
---	-------	------	-------	---

Term	14429	14066	-	0
------	-------	-------	---	---

Intr	14686	14527	-	0
------	-------	-------	---	---

10	Intr	14858	14767	-	0
----	------	-------	-------	---	---

Intr	15138	14935	-	0
------	-------	-------	---	---

Intr	15415	15221	-	0
------	-------	-------	---	---

Intr	15543	15495	-	0
------	-------	-------	---	---

Intr	15749	15646	-	0
------	-------	-------	---	---

15	Intr	15877	15838	-	0
----	------	-------	-------	---	---

Init	16437	16286	-	0
------	-------	-------	---	---

>3702722 /14234

20	len =	1597	nex =	5
----	-------	------	-------	---

Term	50617	50334	-	0
------	-------	-------	---	---

Intr	50852	50691	-	0
------	-------	-------	---	---

Intr	51106	51000	-	0
------	-------	-------	---	---

25	Intr	51288	51207	-	0
----	------	-------	-------	---	---

Init	51930	51764	-	0
------	-------	-------	---	---

>3702722 /35228

30	len =	2920	nex =	6
----	-------	------	-------	---

Term	62691	62388	-	0
------	-------	-------	---	---

Intr	63047	62898	-	0
------	-------	-------	---	---

Intr	63282	63083	-	0
------	-------	-------	---	---

35	Intr	63460	63392	-	0
----	------	-------	-------	---	---

Intr	63685	63542	-	0
------	-------	-------	---	---

Init	64731	64543	-	0
------	-------	-------	---	---

>3702722 /150222

40	len =	885	nex =	2
----	-------	-----	-------	---

Term	67157	66736	-	0
------	-------	-------	---	---

Init	67620	67353	-	0
------	-------	-------	---	---

45				
----	--	--	--	--

>3702723 /639

len =	2568	nex =	11
-------	------	-------	----

50	Term	33692	33399	-	0
----	------	-------	-------	---	---

Intr	33810	33769	-	0
------	-------	-------	---	---

Intr	34153	34055	-	0
------	-------	-------	---	---

Intr	34894	34809	-	0
------	-------	-------	---	---

Intr	35061	34977	-	0
------	-------	-------	---	---

55	Intr	35224	35141	-	0
----	------	-------	-------	---	---

Intr	35373	35294	-	0
------	-------	-------	---	---

Intr	35512	35449	-	0
------	-------	-------	---	---

Intr	35642	35574	-	0
------	-------	-------	---	---

Intr	35787	35707	-	0
------	-------	-------	---	---

60	Init	35966	35863	-	0
----	------	-------	-------	---	---

1231

>3702723 /12930

5	len =	673	nex =	2	
	Init	9767	9878	+	0
	Term	10190	10439	+	0

>3702724 /20287

10	len =	1517	nex =	6	
	Init	20008	20168	+	0
	Intr	20256	20311	+	0
15	Intr	20582	20705	+	0
	Intr	20799	20905	+	0
	Intr	21133	21196	+	0
	Term	21314	21524	+	0

>3702724 /94685

20	len =	259	nex =	1	
	Sngl	38461	38203	-	0

>3702724 /33455

25	len =	1271	nex =	1	
30	Sngl	39470	38200	-	0

>3702724 /6342

35	len =	2050	nex =	5	
	Init	47979	48280	+	0
	Intr	48382	48502	+	0
	Intr	48605	48696	+	0
	Intr	49127	49223	+	0
40	Term	49436	50023	+	0

>3702724 /16773

45	len =	1249	nex =	4	
	Init	64138	64368	+	0
	Intr	64447	64502	+	0
	Intr	64842	64905	+	0
	Term	65169	65386	+	0

>3702724 /95513

50	len =	1636	nex =	3	
55	Term	74797	74565	-	0
	Intr	75178	74949	-	0
	Init	75971	75545	-	0

>3702724 /6984

60

1232

	len =	3134	nex =	8	
	Init	8892	9148	+	0
	Intr	9352	9461	+	0
5	Intr	9555	9825	+	0
	Intr	9920	10055	+	0
	Intr	10237	10603	+	0
	Intr	10691	11059	+	0
	Intr	11151	11229	+	0
10	Term	11334	11533	+	0
>3702728		/207075			
	len =	1647	nex =	8	
15	Term	10490	10127	-	0
	Intr	10656	10573	-	0
	Intr	10792	10748	-	0
	Intr	10942	10888	-	0
20	Intr	11147	11028	-	0
	Intr	11275	11224	-	0
	Intr	11456	11374	-	0
	Init	11773	11604	-	0
25	>3702728	/5200			
	len =	750	nex =	1	
	Sngl	25422	24673	-	0
30	>3702728	/29189			
	len =	578	nex =	1	
35	Sngl	34532	35109	+	0
	>3702728	/25501			
	len =	939	nex =	2	
40	Init	39949	40151	+	0
	Term	40345	40887	+	0
	>3702729	/7579			
45	len =	1270	nex =	1	
	Sngl	28996	29292	+	0
50	>3702730	/18256			
	len =	614	nex =	1	
	Sngl	24396	23783	-	0
55	>3702731	/37901			
	len =	1750	nex =	2	
60	Init	15709	16151	+	0

2025 RELEASE UNDER E.O. 14176

1233

	Term	17205	17451	+	0
	>3702731	/142174			
5	len =	281	nex =	1	
	Sngl	15712	15992	+	0
	>3702731	/205672			
10	len =	617	nex =	1	
	Sngl	1863	1247	-	0
	>3702731	/126555			
15	len =	1884	nex =	2	
	Term	22481	22203	-	0
20	Init	22709	22575	-	0
	>3702731	/36901			
25	len =	2209	nex =	8	
	Term	4259	3981	-	0
	Intr	4544	4368	-	0
	Intr	4802	4628	-	0
30	Intr	4997	4887	-	0
	Intr	5223	5148	-	0
	Intr	5410	5337	-	0
	Intr	5696	5493	-	0
	Init	6189	6014	-	0
	>3702731	/36963			
35	len =	2086	nex =	6	
	Init	7741	8031	+	0
40	Intr	8293	8355	+	0
	Intr	8442	8558	+	0
	Intr	8650	9091	+	0
	Intr	9265	9416	+	0
45	Term	9532	9826	+	0
	>3702731	/27099			
	len =	2791	nex =	7	
50	Init	81625	81712	+	0
	Intr	82523	82571	+	0
	Intr	82720	82828	+	0
	Intr	83013	83106	+	0
	Intr	83540	83673	+	0
55	Intr	83776	83939	+	0
	Term	84172	84415	+	0
	>3702732	/98232			
60	len =	2066	nex =	2	

1234

	Init	13407	13675	+	0
	Term	14812	15472	+	0
5	>3702732 /19319				
	len =	1177	nex =	2	
	Init	18687	18864	+	0
10	Term	19260	19863	+	0
	>3702732 /7700				
	len =	1035	nex =	2	
15	Init	30814	31011	+	0
	Term	31131	31848	+	0
	>3702732 /24349				
20	len =	763	nex =	1	
	Sngl	59677	60439	+	0
25	>3702732 /23018				
	len =	2089	nex =	8	
	Term	60763	60548	-	0
30	Intr	60927	60880	-	0
	Intr	61084	61010	-	0
	Intr	61257	61192	-	0
	Intr	61475	61386	-	0
	Intr	61799	61590	-	0
35	Intr	61920	61887	-	0
	Init	62226	62013	-	0
	>3702732 /41129				
40	len =	474	nex =	1	
	Sngl	6414	5941	-	0
	>3702733 /6599				
45	len =	2383	nex =	9	
	Init	11798	12008	+	0
	Intr	12085	12315	+	0
50	Intr	12405	12484	+	0
	Intr	12606	12773	+	0
	Intr	12850	13025	+	0
	Intr	13106	13175	+	0
	Intr	13283	13437	+	0
55	Intr	13530	13613	+	0
	Term	13738	14180	+	0
	>3702733 /35560				
60	len =	1653	nex =	4	

ORIGINAL PAGE

1235

	Init	24950	25031	+	0
	Intr	25138	25259	+	0
	Intr	25830	25907	+	0
5	Term	25996	26115	+	0
>3702733 /789					
	len =	1531	nex =	1	
10	Sngl	4266	2736	-	0
>3702734 /19116					
15	len =	2455	nex =	9	
	Term	9390	9018	-	0
	Intr	9751	9471	-	0
	Intr	9915	9844	-	0
20	Intr	10035	10000	-	0
	Intr	10280	10119	-	0
	Intr	10491	10387	-	0
	Intr	10697	10611	-	0
	Intr	11111	10989	-	0
25	Init	11472	11255	-	0
>3702734 /7232					
	len =	146	nex =	1	
30	Sngl	3094	3239	+	0
>3702734 /121024					
35	len =	1737	nex =	4	
	Init	42608	42742	+	0
	Intr	43086	43196	+	0
	Intr	43298	43408	+	0
40	Term	43484	43846	+	0
>3702734 /10658					
	len =	579	nex =	1	
45	Sngl	4938	5516	+	0
>3702734 /14354					
50	len =	1181	nex =	2	
	Init	662	1137	+	0
	Term	1645	1842	+	0
55	>3702734 /39557				
	len =	790	nex =	1	
60	Sngl	7453	8241	+	0

2025 RELEASE UNDER E.O. 14176

1236

>3702735 /14236

len = 2126 nex = 7

5	Term	9857	9648	-	0
	Intr	10069	10001	-	0
	Intr	10631	10500	-	0
	Intr	10849	10734	-	0
	Intr	11016	10935	-	0
10	Intr	11194	11124	-	0
	Init	11773	11547	-	0

>3702735 /29871

15 len = 2630 nex = 7

	Term	9857	9489	-	0
	Intr	10069	10001	-	0
	Intr	10631	10500	-	0
20	Intr	10849	10734	-	0
	Intr	11016	10935	-	0
	Intr	11194	11124	-	0
	Init	12118	11803	-	0

25 >3702735 /30034

len = 1860 nex = 5

	Term	12882	12590	-	0
30	Intr	13611	13478	-	0
	Intr	13786	13700	-	0
	Intr	14287	13958	-	0
	Init	14449	14377	-	0

35 >3702735 /1492

len = 2013 nex = 7

	Init	15808	15880	+	0
40	Intr	16103	16150	+	0
	Intr	16235	16282	+	0
	Intr	16379	16450	+	0
	Intr	16538	16693	+	0
	Intr	16788	16891	+	0
45	Term	16967	17259	+	0

>3702735 /39954

len = 2145 nex = 7

50	Term	17743	17402	-	0
	Intr	17875	17828	-	0
	Intr	18119	17961	-	0
	Intr	18361	18220	-	0
55	Intr	18509	18444	-	0
	Intr	18923	18605	-	0
	Init	19546	19195	-	0

>3702735 /7451

60

1237

	len =	288	nex =	1	
	Sngl	19551	19264	-	0
5	>3702735 /19104				
	len =	2153	nex =	7	
	Term	17743	17402	-	0
10	Intr	17875	17828	-	0
	Intr	18119	17961	-	0
	Intr	18361	18220	-	0
	Intr	18509	18444	-	0
	Intr	18923	18605	-	0
15	Init	19554	19195	-	0
	>3702735 /37772				
	len =	2186	nex =	7	
20	Term	17743	17415	-	0
	Intr	17875	17828	-	0
	Intr	18119	17961	-	0
	Intr	18361	18220	-	0
25	Intr	18509	18444	-	0
	Intr	18923	18605	-	0
	Init	19600	19195	-	0
	>3702735 /108736				
30	len =	2183	nex =	7	
	Term	17743	17418	-	0
	Intr	17875	17828	-	0
35	Intr	18119	17961	-	0
	Intr	18361	18220	-	0
	Intr	18509	18444	-	0
	Intr	18923	18605	-	0
	Init	19600	19195	-	0
40	>3702735 /3012				
	len =	970	nex =	1	
45	Sngl	21165	20204	-	0
	>3702735 /16025				
	len =	1114	nex =	1	
50	Sngl	28361	29474	+	0
	>3702735 /40827				
55	len =	1870	nex =	7	
	Init	32059	32271	+	0
	Intr	32368	32648	+	0
	Intr	32727	32792	+	0
60	Intr	32884	33024	+	0

2025 RELEASE UNDER E.O. 14176

1238

Intr	33122	33250	+	0
Intr	33351	33494	+	0
Term	33582	33926	+	0

5 >3702735 /38864

```
len = 2110    nex = 3
```

	Init	50260	50547	+	0
10	Intr	50645	50722	+	0
	Term	50817	51231	+	0

>3702735 /1037

```
15      len =      1950      nex =          4
```

Init	52134	52251	+	0
Intr	52369	52437	+	0
Intr	52728	52834	+	0
Term	52940	53499	+	0

>3702735 /22308

```
len =      408      nex =      1
```

Sngl	53268	53675	+	0
------	-------	-------	---	---

>3702735 /13359

```
30      len =      1750      nex =          7
```

Init	61031	61277	+	0
Intr	61431	61563	+	0
Intr	61649	61765	+	0
Intr	61882	61947	+	0
Intr	62033	62071	+	0
Intr	62165	62287	+	0
Term	62370	62773	+	0

40 >3702735 /26013

```
len = 1510    nex = 4
```

	Term	65133	64653	-	0
45	Intr	65309	65230	-	0
	Intr	65703	65393	-	0
	Init	66162	65839	-	0

>3702736 /21033

```

50      len =      685      nex =      1

```

Sngl	14295	13611	-	0
------	-------	-------	---	---

55 >3702736 /22417

```
len = 2244 nex = 9
```

	Init	22337	22446	+	0
60	Intr	22746	22848	+	0

1239

	Intr	22935	23021	+	0
	Intr	23099	23226	+	0
	Intr	23309	23405	+	0
	Intr	23513	23590	+	0
5	Intr	23673	23898	+	0
	Intr	23978	24201	+	0
	Term	24299	24580	+	0
	>3702736	/6639			
10	len =	1465	nex =	2	
	Init	2970	3755	+	0
	Term	3857	4434	+	0
15	>3702736	/150178			
	len =	495	nex =	1	
20	Sngl	41469	41963	+	0
	>3702736	/99738			
	len =	551	nex =	1	
25	Sngl	48180	48730	+	0
	>3702737	/35235			
30	len =	2470	nex =	12	
	Term	23292	23060	-	0
	Intr	23486	23403	-	0
	Intr	23671	23567	-	0
35	Intr	23913	23788	-	0
	Intr	24083	24012	-	0
	Intr	24325	24250	-	0
	Intr	24509	24468	-	0
	Intr	24649	24594	-	0
40	Intr	24929	24816	-	0
	Intr	25118	25065	-	0
	Intr	25399	25274	-	0
	Init	25529	25486	-	0
45	>3702737	/18419			
	len =	2968	nex =	13	
	Term	23292	23058	-	0
50	Intr	23486	23403	-	0
	Intr	23671	23567	-	0
	Intr	23913	23788	-	0
	Intr	24083	24012	-	0
	Intr	24325	24250	-	0
55	Intr	24509	24468	-	0
	Intr	24649	24594	-	0
	Intr	24929	24816	-	0
	Intr	25118	25065	-	0
	Intr	25399	25274	-	0
60	Intr	25622	25486	-	0

1240

	Init	26025	25713	-	0
	>3702737 /41468				
5	len =	1214	nex =	5	
	Init	61712	61761	+	0
	Intr	61844	62014	+	0
	Intr	62113	62182	+	0
10	Intr	62277	62397	+	0
	Term	62489	62925	+	0
	>3702737 /25523				
15	len =	1270	nex =	2	
	Term	9043	8460	-	0
	Init	9725	9132	-	0
20	>3702738 /3619				
	len =	1734	nex =	3	
	Init	8288	8505	+	0
25	Intr	9426	9614	+	0
	Term	9689	10021	+	0
	>3702739 /38027				
30	len =	342	nex =	1	
	Sngl	342	1	-	0
	>3702739 /11852				
35	len =	1734	nex =	1	
	Sngl	83965	83716	-	0
40	>3738088 /29306				
	len =	2416	nex =	9	
	Term	21390	21051	-	0
45	Intr	21586	21492	-	0
	Intr	21773	21684	-	0
	Intr	21894	21859	-	0
	Intr	22056	21973	-	0
	Intr	22222	22133	-	0
50	Intr	22628	22314	-	0
	Intr	23079	22864	-	0
	Init	23466	23181	-	0
	>3738275 /14182				
55	len =	975	nex =	2	
	Init	18779	18953	+	0
60	Term	19387	19753	+	0

1241

>3738275 /8386

len = 1164 nex = 4

5	Term	22568	22293	-	0
	Intr	22774	22653	-	0
	Intr	23106	23012	-	0
	Init	23456	23209	-	0

10 >3738275 /39367

len = 1990 nex = 8

	Term	38894	38400	-	0
15	Intr	39077	38993	-	0
	Intr	39243	39152	-	0
	Intr	39444	39334	-	0
	Intr	39608	39534	-	0
	Intr	39848	39753	-	0
20	Intr	40152	39939	-	0
	Init	40384	40234	-	0

>3738275 /21006

25	len =	619	nex =	1	
	Sngl	72476	73094	+	0

>3738275 /39915

30	len =	1487	nex =	4	
	Init	75590	76151	+	0
	Intr	76230	76391	+	0
35	Intr	76475	76746	+	0
	Term	76819	77076	+	0

>3738275 /10341

40	len =	929	nex =	2	
	Init	82166	82372	+	0
	Term	82451	83094	+	0

45 >3738275 /5414

	len =	699	nex =	1	
	Sngl	85117	84419	-	0

50 >3738275 /31153

	len =	587	nex =	1	
55	Sngl	91661	92247	+	0

>3738275 /7887

60	len =	1700	nex =	2	
----	-------	------	-------	---	--

20250901 09:01:00

1242

	Init	96923	97953	+	0
	Term	98161	98622	+	0
	>3738313 /11227				
5	len =	1048	nex =	1	
	Sngl	31224	30177	-	0
10	>3738313 /14133				
	len =	2123	nex =	6	
	Init	40926	41117	+	0
15	Intr	41351	41508	+	0
	Intr	41614	41805	+	0
	Intr	42263	42416	+	0
	Intr	42517	42709	+	0
	Term	42788	43048	+	0
20	>3738313 /4850				
	len =	976	nex =	4	
25	Init	45374	45479	+	0
	Intr	45584	45606	+	0
	Intr	45853	45974	+	0
	Term	46056	46332	+	0
30	>3738313 /13580				
	len =	1060	nex =	5	
	Init	47214	47374	+	0
35	Intr	47473	47666	+	0
	Intr	47779	47863	+	0
	Intr	47946	48011	+	0
	Term	48097	48273	+	0
40	>3738313 /28044				
	len =	3250	nex =	11	
	Term	50661	50155	-	0
45	Intr	51114	50735	-	0
	Intr	51543	51199	-	0
	Intr	51756	51628	-	0
	Intr	51950	51850	-	0
	Intr	52089	52030	-	0
50	Intr	52236	52165	-	0
	Intr	52471	52328	-	0
	Intr	52634	52563	-	0
	Intr	52848	52716	-	0
	Init	53400	52959	-	0
55	>3738313 /11327				
	len =	2590	nex =	6	
60	Term	65078	64374	-	0

1243

	Intr	65397	65161	-	0
	Intr	65607	65496	-	0
	Intr	66003	65690	-	0
	Intr	66278	66114	-	0
5	Init	66956	66369	-	0
	>3738313		/41558		
	len =	2170	nex =	9	
10	Init	92485	92831	+	0
	Intr	92913	92952	+	0
	Intr	93039	93098	+	0
	Intr	93463	93515	+	0
15	Intr	93603	93652	+	0
	Intr	93744	93777	+	0
	Intr	93875	93947	+	0
	Intr	94031	94123	+	0
	Term	94234	94650	+	0
20	>3746057		/21261		
	len =	1285	nex =	1	
25	Sngl	15904	14620	-	0
	>3746057		/36533		
	len =	3692	nex =	5	
30	Init	27102	27418	+	0
	Intr	27863	28034	+	0
	Intr	29868	30043	+	0
	Intr	30142	30300	+	0
35	Term	30381	30793	+	0
	>3746057		/10986		
	len =	574	nex =	2	
40	Term	32722	32483	-	0
	Init	33056	32809	-	0
	>3757512		/34622		
	len =	922	nex =	3	
	Term	11574	11216	-	0
	Intr	11801	11665	-	0
50	Init	12137	11918	-	0
	>3757512		/156725		
	len =	790	nex =	2	
55	Term	13186	13046	-	0
	Init	13828	13761	-	0
	>3757512		/36208		
60					

1244

	len =	2715	nex =	5	
	Term	11574	11127	-	0
5	Intr	11801	11665	-	0
	Intr	12158	11918	-	0
	Intr	12416	12250	-	0
	Init	13186	13044	-	0
	>3757512 /17067				
10	len =	1177	nex =	4	
	Init	48329	48418	+	0
	Intr	48697	48796	+	0
15	Intr	48877	49001	+	0
	Term	49084	49505	+	0
	>3757512 /1629				
20	len =	1161	nex =	3	
	Init	48697	48796	+	0
	Intr	48877	49001	+	0
	Term	49084	49491	+	0
25	>3757512 /34897				
	len =	1056	nex =	3	
30	Init	48697	48796	+	0
	Intr	48877	49001	+	0
	Term	49084	49445	+	0
	>3757512 /103391				
35	len =	807	nex =	3	
	Init	48697	48796	+	0
	Intr	48877	49001	+	0
40	Term	49084	49503	+	0
	>3757512 /10206				
45	len =	730	nex =	3	
	Term	50124	49809	-	0
	Intr	50347	50233	-	0
	Init	50535	50445	-	0
50	>3757512 /16224				
	len =	4768	nex =	12	
	Term	50124	49849	-	0
55	Intr	50347	50233	-	0
	Intr	50533	50445	-	0
	Intr	51384	51295	-	0
	Intr	51786	51712	-	0
	Intr	51965	51882	-	0
60	Intr	52405	52174	-	0

1245

	Intr	52663	52608	-	0
	Intr	52821	52765	-	0
	Intr	52984	52903	-	0
	Intr	54261	53779	-	0
5	Init	54612	54457	-	0
	>3763915 /39279				
	len =	1270	nex =	2	
10	Term	19272	18719	-	0
	Init	19986	19634	-	0
	>3763915 /20820				
15	len =	2055	nex =	5	
	Init	45580	45662	+	0
	Intr	45753	46015	+	0
20	Intr	46153	46221	+	0
	Intr	46358	46416	+	0
	Term	46586	46687	+	0
	>3763915 /12017				
25	len =	2775	nex =	7	
	Init	50003	50101	+	0
	Intr	50315	50367	+	0
30	Intr	50890	50957	+	0
	Intr	51035	51253	+	0
	Intr	51335	51863	+	0
	Intr	52052	52294	+	0
	Term	52385	52777	+	0
35	>3763915 /34599				
	len =	1286	nex =	5	
40	Init	56203	56557	+	0
	Intr	56631	56904	+	0
	Intr	56983	57115	+	0
	Intr	57204	57290	+	0
	Term	57378	57488	+	0
45	>3763915 /311				
	len =	4030	nex =	13	
50	Term	60393	60334	-	0
	Intr	60545	60500	-	0
	Intr	60724	60645	-	0
	Intr	60942	60809	-	0
	Intr	61216	61109	-	0
55	Intr	61408	61339	-	0
	Intr	61923	61846	-	0
	Intr	62172	62050	-	0
	Intr	62408	62304	-	0
	Intr	63134	63057	-	0
60	Intr	63482	63426	-	0

2025-08-08 09:29:29

1246

Intr	63977	63861	-	0
Init	64363	64190	-	0

>3763915 /31507

5	len =	1402	nex =	2
---	-------	------	-------	---

Init	7990	8335	+	0
Term	9245	9391	+	0

10	>3763915	/18447
----	----------	--------

len =	1714	nex =	3
-------	------	-------	---

15	Term	87259	87159	-	0
	Intr	87414	87348	-	0
	Init	87961	87899	-	0

>3763944 /8742

20	len =	1230	nex =	4
----	-------	------	-------	---

Init	16958	17118	+	0
Intr	17220	17330	+	0

25	Intr	17589	17767	+	0
	Term	17874	18187	+	0

>3763944 /36974

30	len =	1990	nex =	5
----	-------	------	-------	---

Init	1865	2006	+	0
Intr	2096	2209	+	0

35	Intr	2304	2531	+	0
	Intr	2952	3391	+	0

Term	3484	3851	+	0
------	------	------	---	---

>3763944 /8686

40	len =	1578	nex =	6
----	-------	------	-------	---

Init	19154	19555	+	0
Intr	19635	19679	+	0

45	Intr	19765	19837	+	0
	Intr	19939	20036	+	0

Term	20136	20206	+	0
	20399	20731	+	0

>3763944 /19199

50	len =	1228	nex =	6
----	-------	------	-------	---

Init	19315	19555	+	0
Intr	19635	19679	+	0

55	Intr	19765	19837	+	0
	Intr	19939	20036	+	0

Intr	20136	20206	+	0
Term	20399	20542	+	0

60	>3763944	/9640
----	----------	-------

1247

	len =	772	nex =	2	
	Init	3064	3391	+	0
5	Term	3484	3835	+	0
>3763944 /40008					
	len =	2252	nex =	7	
10	Term	33666	33350	-	0
	Intr	34075	33766	-	0
	Intr	34206	34161	-	0
	Intr	34967	34896	-	0
15	Intr	35185	35059	-	0
	Intr	35397	35293	-	0
	Init	35601	35504	-	0
>3763944 /3321					
20	len =	1486	nex =	6	
	Term	48162	48090	-	0
	Intr	48339	48287	-	0
25	Intr	48651	48603	-	0
	Intr	48901	48837	-	0
	Intr	49077	48990	-	0
	Init	49575	49178	-	0
30	>3763944	/35300			
	len =	339	nex =	1	
	Sngl	54201	53863	-	0
35	>3763944	/37313			
	len =	2087	nex =	6	
40	Init	65398	65640	+	0
	Intr	65729	65831	+	0
	Intr	66198	66283	+	0
	Intr	66459	66697	+	0
	Intr	66803	66992	+	0
45	Term	67215	67484	+	0
>3763944 /16451					
	len =	1295	nex =	3	
50	Term	78903	78431	-	0
	Intr	79217	79086	-	0
	Init	79720	79611	-	0
55	>3763944	/8049			
	len =	1390	nex =	3	
	Term	78903	78426	-	0
60	Intr	79217	79086	-	0

00190-022520

1248

	Init	79807	79611	-	0
	>3763944	/43034			
5	len =	386	nex =	1	
	Sngl	79807	79611	-	0
	>3766106	/122624			
10	len =	190	nex =	1	
	Sngl	13182	12996	-	0
15	>3766106	/17815			
	len =	4706	nex =	19	
	Term	13428	12997	-	0
20	Intr	13611	13523	-	0
	Intr	13786	13708	-	0
	Intr	13982	13885	-	0
	Intr	14288	14184	-	0
	Intr	14458	14372	-	0
25	Intr	14840	14646	-	0
	Intr	15050	14944	-	0
	Intr	15272	15140	-	0
	Intr	15413	15351	-	0
	Intr	15600	15499	-	0
30	Intr	15756	15685	-	0
	Intr	15905	15849	-	0
	Intr	16194	16093	-	0
	Intr	16361	16324	-	0
	Intr	16560	16449	-	0
35	Intr	16723	16667	-	0
	Intr	16859	16804	-	0
	Init	17096	16970	-	0
	>3766106	/31830			
40	len =	1468	nex =	1	
	Sngl	21312	20870	-	0
45	>3766106	/2763			
	len =	1609	nex =	7	
	Term	25723	25696	-	0
50	Intr	25940	25843	-	0
	Intr	26117	26044	-	0
	Intr	26241	26200	-	0
	Intr	26573	26515	-	0
	Intr	26960	26874	-	0
55	Init	27166	27008	-	0
	>3766106	/19071			
	len =	1931	nex =	8	
60					

2025 RELEASE UNDER E.O. 14176

1249

	Term	25723	25503	-	0
	Intr	25940	25843	-	0
	Intr	26117	26044	-	0
	Intr	26241	26200	-	0
5	Intr	26573	26515	-	0
	Intr	26960	26874	-	0
	Intr	27166	27139	-	0
	Init	27433	27265	-	0
10	>3766106 /772				
	len =	1100	nex =	3	
	Init	55679	55811	+	0
15	Intr	55892	56190	+	0
	Term	56287	56543	+	0
	>3766106 /42210				
20	len =	1404	nex =	4	
	Term	4966	4948	-	0
	Intr	5280	5077	-	0
	Intr	5648	5451	-	0
25	Init	6351	6198	-	0
	>3766106 /37467				
	len =	3610	nex =	15	
30	Init	6698	6918	+	0
	Intr	7000	7028	+	0
	Intr	7114	7151	+	0
	Intr	7313	7400	+	0
35	Intr	7670	7775	+	0
	Intr	7890	7960	+	0
	Intr	8194	8263	+	0
	Intr	8365	8427	+	0
	Intr	8718	8789	+	0
40	Intr	8898	8999	+	0
	Intr	9219	9293	+	0
	Intr	9415	9513	+	0
	Intr	9747	9863	+	0
	Intr	9947	10003	+	0
45	Term	10101	10306	+	0
	>3766106 /31765				
	len =	1407	nex =	5	
50	Term	82104	82012	-	0
	Intr	82327	82239	-	0
	Intr	82516	82454	-	0
	Intr	82689	82600	-	0
55	Init	82977	82917	-	0
	>3766106 /32117				
	len =	2329	nex =	9	
60					

1250

	Term	81812	81701	-	0
	Intr	82104	82012	-	0
	Intr	82327	82239	-	0
5	Intr	82516	82454	-	0
	Intr	82689	82600	-	0
	Intr	82977	82917	-	0
	Intr	83164	83070	-	0
	Intr	83370	83281	-	0
10	Init	84029	83633	-	0
	>3779020	/7600			
	len =	928	nex =	3	
15	Term	43560	43284	-	0
	Intr	43921	43879	-	0
	Init	44211	44021	-	0
	>3785968	/41002			
20	len =	2151	nex =	9	
	Init	47820	47930	+	0
	Intr	48026	48147	+	0
25	Intr	48227	48287	+	0
	Intr	48637	48705	+	0
	Intr	48785	48903	+	0
	Intr	49026	49173	+	0
	Intr	49271	49415	+	0
30	Intr	49514	49600	+	0
	Term	49702	49962	+	0
	>3785968	/18820			
35	len =	2037	nex =	4	
	Term	51353	50716	-	0
	Intr	51853	51786	-	0
	Intr	52406	52062	-	0
40	Init	52752	52479	-	0
	>3785968	/11949			
	len =	1210	nex =	1	
45	Sngl	56312	55108	-	0
	>3785968	/35997			
50	len =	1090	nex =	2	
	Init	62489	62738	+	0
	Term	63173	63578	+	0
55	>3785968	/17603			
	len =	1210	nex =	2	
	Init	62489	62738	+	0
60	Term	63173	63691	+	0

2025 RELEASE UNDER E.O. 14176

1251

>3785968 /2422

5	len =	956	nex =	2	
	Init	832	994	+	0
	Term	1320	1787	+	0

>3785968 /32861

10	len =	1930	nex =	5	
	Init	97394	97474	+	0
	Intr	97567	97743	+	0
15	Intr	97843	97921	+	0
	Intr	98036	98151	+	0
	Term	98536	99144	+	0

>3785992 /3294

20	len =	1600	nex =	6	
	Term	14853	14529	-	0
	Intr	15039	14947	-	0
25	Intr	15343	15131	-	0
	Intr	15667	15434	-	0
	Intr	15888	15743	-	0
	Init	16128	15979	-	0

>3785992 /32721

30	len =	2131	nex =	6	
	Init	18649	18757	+	0
35	Intr	18865	19046	+	0
	Intr	19570	19788	+	0
	Intr	19870	20091	+	0
	Intr	20183	20254	+	0
40	Term	20343	20779	+	0

>3785992 /40283

	len =	1785	nex =	4	
45	Init	21029	21179	+	0
	Intr	21550	21664	+	0
	Intr	21810	21961	+	0
	Term	22069	22813	+	0

>3785992 /17861

50	len =	2802	nex =	5	
	Init	30317	30479	+	0
55	Intr	30906	31025	+	0
	Intr	31867	31981	+	0
	Intr	32237	32330	+	0
	Term	32412	32768	+	0

>3785992 /35493

1252

	len =	3170	nex =	9	
	Init	33260	33564	+	0
5	Intr	33721	33919	+	0
	Intr	34406	34469	+	0
	Intr	34703	34847	+	0
	Intr	35156	35242	+	0
	Intr	35398	35555	+	0
10	Intr	35655	35703	+	0
	Intr	35816	35940	+	0
	Term	36021	36429	+	0
	>3785992	/37377			
15	len =	2858	nex =	10	
	Init	33355	33564	+	0
	Intr	33721	33919	+	0
20	Intr	34406	34469	+	0
	Intr	34703	34847	+	0
	Intr	35156	35242	+	0
	Intr	35398	35555	+	0
	Intr	35655	35703	+	0
25	Intr	35816	35940	+	0
	Intr	36021	36123	+	0
	Term	36144	36212	+	0
	>3785992	/21746			
30	len =	2965	nex =	9	
	Term	45531	45288	-	0
	Intr	45743	45645	-	0
35	Intr	45891	45826	-	0
	Intr	46097	45977	-	0
	Intr	46418	46351	-	0
	Intr	46701	46630	-	0
	Intr	47414	47230	-	0
40	Intr	47749	47514	-	0
	Init	48252	47839	-	0
	>3785992	/101298			
45	len =	250	nex =	1	
	Sngl	82983	82739	-	0
	>3785992	/25626			
50	len =	2398	nex =	6	
	Term	83028	82729	-	0
	Intr	83330	83293	-	0
55	Intr	83469	83420	-	0
	Intr	83997	83830	-	0
	Intr	84473	84083	-	0
	Init	85126	84720	-	0
60	>3785992	/18328			

2025 RELEASE UNDER E.O. 14176

1254

Intr	45268	45351	+	0
Intr	45442	45609	+	0
Term	45706	45953	+	0

5 >3805839 /5482

```
len = 3370    nex = 12
```

10	Init	47368	47506	+	0
	Intr	47679	47841	+	0
	Intr	48042	48107	+	0
	Intr	48209	48294	+	0
	Intr	48479	48596	+	0
	Intr	48680	48886	+	0
15	Intr	49008	49084	+	0
	Intr	49250	49343	+	0
	Intr	49606	49671	+	0
	Intr	49774	49851	+	0
	Intr	49955	50056	+	0
20	Term	50145	50455	+	0

>3805839 /17725

```

25      len =      2056      nex =          6

```

Term	50772	50483	-	0
Intr	50949	50881	-	0
Intr	51247	51043	-	0
Intr	51692	51610	-	0
Intr	52241	52102	-	0
Init	52538	52356	-	0

>3805839 /1095

```
35      len =      1127      nex =          3
```

Term	70923	70590	-	0
Intr	71436	71382	-	0
Init	71716	71529	-	0

>3805839 /8360

```
len = 1095    nex = 3
```

45	Term	94509	94155	-	0
	Intr	94708	94595	-	0
	Init	95249	95131	-	0

>3831437 /3743

```
len = 941 nex = 3
```

Init	6216	6447	+	0
Intr	6552	6618	+	0
Term	6704	7156	+	0

>3831448 /9671

```

len = 1169 nex = 2
60

```


Reference No. 2750-942P

1255

	Init	40447	40831	+	0
	Term	41264	41615	+	0
	>3831448	/17467			
5	len =	395	nex =	1	
	Sngl	61543	61937	+	0
10	>3831448	/35104			
	len =	1522	nex =	1	
	Sngl	8526	7008	-	0
15	>3831448	/37237			
	len =	3011	nex =	7	
20	Term	88224	87712	-	0
	Intr	88432	88334	-	0
	Intr	88649	88483	-	0
	Intr	89505	89463	-	0
	Intr	89768	89673	-	0
25	Intr	89984	89868	-	0
	Init	90722	90415	-	0
	>3849811	/9623			
30	len =	1165	nex =	4	
	Term	896	533	-	0
	Intr	1195	997	-	0
	Intr	1462	1307	-	0
35	Init	1697	1546	-	0
	>3849811	/1568			
	len =	1042	nex =	4	
40	Term	23646	23559	-	0
	Intr	23842	23733	-	0
	Intr	24392	24313	-	0
	Init	24600	24516	-	0
45	>3849811	/30470			
	len =	1123	nex =	3	
50	Term	23842	23486	-	0
	Intr	24392	24313	-	0
	Init	24608	24516	-	0
	>3849811	/7182			
55	len =	1193	nex =	4	
	Term	23646	23450	-	0
	Intr	23842	23733	-	0
60	Intr	24392	24313	-	0

2025 RELEASE UNDER E.O. 14176

1256

	Init	24642	24516	-	0
	>3849811 /36468				
5	len =	1339	nex =	1	
	Sngl	28830	27492	-	0
	>3849811 /33391				
10	len =	1311	nex =	3	
	Init	29563	29631	+	0
	Intr	29985	30486	+	0
15	Term	30573	30873	+	0
	>3849811 /35307				
	len =	730	nex =	2	
20	Init	29985	30486	+	0
	Term	30573	30713	+	0
	>3849811 /24946				
25	len =	772	nex =	1	
	Sngl	37425	38196	+	0
	>3849811 /19020				
30	len =	2018	nex =	6	
	Init	51039	51216	+	0
35	Intr	51447	51560	+	0
	Intr	51656	51804	+	0
	Intr	51969	52041	+	0
	Intr	52302	52341	+	0
	Term	52685	52767	+	0
40	>3849811 /18200				
	len =	1310	nex =	2	
45	Init	59367	59612	+	0
	Term	60232	60674	+	0
	>3849811 /12305				
50	len =	512	nex =	1	
	Sngl	5941	6452	+	0
	>3859590 /34747				
55	len =	741	nex =	3	
	Term	39398	39023	-	0
	Intr	39602	39499	-	0
60	Init	39763	39682	-	0

1257

>3859590 /36898

5	len =	1476	nex =	2	
	Init	53829	54125	+	0
	Term	54213	54342	+	0

>3859590 /15259

10	len =	343	nex =	1	
	Sngl	58875	59217	+	0

15 >3859590 /20761

	len =	2429	nex =	11	
20	Term	74605	74285	-	0
	Intr	74729	74694	-	0
	Intr	74938	74825	-	0
	Intr	75236	75018	-	0
	Intr	75451	75326	-	0
	Intr	75613	75539	-	0
25	Intr	75707	75703	-	0
	Intr	75970	75814	-	0
	Intr	76141	76061	-	0
	Intr	76278	76225	-	0
	Init	76713	76474	-	0

>3859590 /95621

	len =	1150	nex =	4	
35	Term	77280	77093	-	0
	Intr	77795	77518	-	0
	Intr	77994	77893	-	0
	Init	78208	78069	-	0

40 >3859590 /15383

	len =	1038	nex =	3	
45	Term	84801	84590	-	0
	Intr	85252	84874	-	0
	Init	85627	85486	-	0

>3859590 /29072

50	len =	712	nex =	1	
	Sngl	88825	89536	+	0

>3859658 /17194

55	len =	515	nex =	1	
	Sngl	16172	16686	+	0

60 >3859658 /107993

1258

	len =	586	nex =	1	
5	Sngl	16174	16759	+	0
	>3859658	/33603			
	len =	409	nex =	1	
10	Sngl	16552	16144	-	0
	>3859658	/6589			
15	len =	2209	nex =	9	
	Init	20101	20257	+	0
	Intr	20354	20412	+	0
	Intr	20723	20875	+	0
	Intr	20969	21017	+	0
20	Intr	21116	21241	+	0
	Intr	21326	21379	+	0
	Intr	21582	21664	+	0
	Intr	21751	21856	+	0
	Term	21928	22293	+	0
25	>3859658	/29040			
	len =	744	nex =	3	
30	Init	21625	21664	+	0
	Intr	21751	21856	+	0
	Term	21928	22368	+	0
	>3859658	/31672			
35	len =	712	nex =	2	
	Term	40705	40329	-	0
40	Init	41020	40903	-	0
	>3859658	/109432			
	len =	610	nex =	2	
45	Term	48314	48122	-	0
	Init	48723	48437	-	0
	>3859658	/38416			
50	len =	403	nex =	1	
	Sngl	49905	50307	+	0
	>3859658	/29886			
55	len =	2037	nex =	5	
	Term	50573	50297	-	0
	Intr	50791	50663	-	0
60	Intr	51573	51274	-	0

0050326-061600

1259

	Intr	51886	51658	-	0
	Init	52333	52049	-	0
	>3859658		/38404		
5	len =	2978	nex =	10	
	Term	52912	52479	-	0
	Intr	53209	53003	-	0
10	Intr	53605	53285	-	0
	Intr	53757	53707	-	0
	Intr	53947	53843	-	0
	Intr	54222	54160	-	0
	Intr	54623	54507	-	0
15	Intr	54886	54819	-	0
	Intr	55081	55005	-	0
	Init	55456	55320	-	0
	>3859658		/5743		
20	len =	790	nex =	1	
	Sngl	63524	64304	+	0
	>3859658		/6636		
25	len =	3730	nex =	9	
	Term	85246	84985	-	0
30	Intr	85487	85360	-	0
	Intr	85826	85560	-	0
	Intr	85995	85914	-	0
	Intr	86282	86121	-	0
	Intr	86558	86366	-	0
35	Intr	87259	87159	-	0
	Intr	87578	87376	-	0
	Init	88293	88190	-	0
	>3859658		/6375		
40	len =	1398	nex =	3	
	Term	91234	91166	-	0
	Intr	91430	91335	-	0
45	Init	92183	91722	-	0
	>3860242		/1860		
	len =	833	nex =	2	
50	Init	19578	19983	+	0
	Term	20109	20390	+	0
	>3860242		/123227		
55	len =	1390	nex =	4	
	Init	22933	23272	+	0
	Intr	23512	23607	+	0
60	Intr	23806	23845	+	0

1260

	Term	24006	24320	+	0
	>3860242	/29893			
5	len =	1317	nex =	4	
	Init	38909	39035	+	0
	Intr	39156	39214	+	0
	Intr	39796	39839	+	0
10	Term	39982	40225	+	0
	>3860242	/40339			
	len =	705	nex =	1	
15	Sngl	45860	45156	-	0
	>3860242	/38650			
20	len =	1396	nex =	1	
	Sngl	46542	45147	-	0
	>3860242	/25214			
25	len =	2937	nex =	4	
	Term	50568	50049	-	0
	Intr	50808	50653	-	0
30	Intr	51048	50902	-	0
	Init	52985	52151	-	0
	>3860242	/21711			
35	len =	2316	nex =	9	
	Init	62073	62233	+	0
	Intr	62317	62379	+	0
	Intr	62477	62543	+	0
40	Intr	62938	62988	+	0
	Intr	63081	63161	+	0
	Intr	63509	63621	+	0
	Intr	63713	63830	+	0
	Intr	63931	64057	+	0
45	Term	64162	64388	+	0
	>3860242	/96159			
	len =	1407	nex =	4	
50	Term	64781	64588	-	0
	Intr	64990	64893	-	0
	Intr	65621	65504	-	0
	Init	65994	65896	-	0
55	>3860243	/14203			
	len =	949	nex =	0	
60	>3860243	/19048			

1261

	len =	1570	nex =	4	
	Init	115309	115528	+	0
5	Intr	115614	115764	+	0
	Intr	115886	116091	+	0
	Term	116181	116399	+	0
	>3860243 /28205				
10	len =	2483	nex =	9	
	Init	12779	13109	+	0
	Intr	13268	13360	+	0
15	Intr	13462	13569	+	0
	Intr	13849	13963	+	0
	Intr	14039	14172	+	0
	Intr	14261	14362	+	0
	Intr	14472	14564	+	0
20	Intr	14654	14789	+	0
	Term	14881	15261	+	0
	>3860243 /6892				
25	len =	2155	nex =	8	
	Init	12829	13109	+	0
	Intr	13268	13360	+	0
	Intr	13462	13569	+	0
30	Intr	13849	13963	+	0
	Intr	14039	14172	+	0
	Intr	14261	14362	+	0
	Intr	14472	14564	+	0
	Term	14654	14983	+	0
35	>3860243 /113089				
	len =	1489	nex =	5	
40	Init	13495	13963	+	0
	Intr	14039	14172	+	0
	Intr	14261	14362	+	0
	Intr	14472	14564	+	0
	Term	14654	14983	+	0
45	>3860243 /23972				
	len =	1889	nex =	7	
50	Term	16019	15795	-	0
	Intr	16256	16115	-	0
	Intr	16406	16348	-	0
	Intr	16535	16500	-	0
	Intr	16972	16620	-	0
55	Intr	17135	17065	-	0
	Init	17683	17580	-	0
	>3860243 /33700				
60	len =	910	nex =	1	

1262

	Sngl	19112	18209	-	0
	>3860243	/37739			
5	len =	1518	nex =	1	
	Sngl	22909	21392	-	0
10	>3860243	/112146			
	len =	1001	nex =	2	
	Init	36410	36592	+	0
15	Term	36690	36790	+	0
	>3860243	/3359			
	len =	2034	nex =	3	
20	Init	58793	58908	+	0
	Intr	59274	59312	+	0
	Term	59418	59832	+	0
25	>3860243	/19986			
	len =	1037	nex =	4	
	Term	60333	60150	-	0
30	Intr	60655	60423	-	0
	Intr	60984	60894	-	0
	Init	61186	61074	-	0
	>3860243	/207594			
35	len =	407	nex =	2	
	Term	62033	61867	-	0
	Init	62273	62139	-	0
40	>3860243	/9845			
	len =	1062	nex =	3	
45	Term	62033	61843	-	0
	Intr	62371	62139	-	0
	Init	62751	62661	-	0
	>3860243	/120911			
50	len =	959	nex =	3	
	Term	62033	61946	-	0
	Intr	62371	62139	-	0
55	Init	62751	62661	-	0
	>3860243	/36076			
60	len =	1039	nex =	3	

[illegible]

Term	62033	61867	-	0
Intr	62371	62139	-	0
Init	62751	62661	-	0

```
len = 2090    nex = 9
```

10	Init	73691	74012	+	0
	Intr	74216	74300	+	0
	Intr	74385	74501	+	0
	Intr	74585	74649	+	0
15	Intr	74744	74832	+	0
	Intr	74933	75012	+	0
	Intr	75094	75205	+	0
	Intr	75307	75426	+	0
	Term	75511	75780	+	0

```
len = 2891 nex = 8
```

	Term	96072	95749	-	0
25	Intr	96363	96208	-	0
	Intr	96564	96466	-	0
	Intr	96729	96655	-	0
	Intr	96877	96809	-	0
	Intr	97159	96961	-	0
	Intr	97879	97774	-	0
30	Init	98639	97974	-	0

>3868722 /125396

```

35      len =      179      nex =      1
      Snq1 14027 13849      -      0

```

>3868723 /35059

```

40      len =      3373      nex =          6
        Init 47360 47692 + 0
        Intr 47785 47840 + 0
        Intr 47939 47993 + 0
45      Intr 49588 49668 + 0
        Intr 49756 49979 + 0
        Term 50524 50732 + 0

```

50 >3868723 /101605

```
len =      415    nex =      1
Snq1 52231    51821    -      0
```

55 >3868723 /42560

```
len = 1118    nex = 2
```

	Term	62440	61893	-	0
60	Init	63010	62468	-	0

1264

>3869062 /15995

5	len =	1300	nex =	2	
	Init	5593	5774	+	0
	Term	5873	6151	+	0

>3869063 /12997

10	len =	1210	nex =	2	
	Init	39503	39620	+	0
	Term	39709	40708	+	0

>3869063 /2393

15	len =	813	nex =	1	
20	Sngl	49706	50518	+	0

>3869063 /37862

25	len =	2178	nex =	7	
	Init	65062	65329	+	0
	Intr	65646	65723	+	0
	Intr	65820	65987	+	0
	Intr	66119	66167	+	0
30	Intr	66275	66348	+	0
	Intr	66564	66657	+	0
	Term	66756	67239	+	0

>3869064 /37349

35	len =	2599	nex =	8	
	Term	1690	525	-	0
	Intr	1966	1776	-	0
40	Intr	2274	2069	-	0
	Intr	2446	2367	-	0
	Intr	2588	2535	-	0
	Intr	2723	2663	-	0
45	Intr	2912	2839	-	0
	Init	3118	2994	-	0

>3869065 /119129

50	len =	1317	nex =	3	
	Term	30028	29747	-	0
	Intr	30375	30116	-	0
	Init	31063	30948	-	0

55 >3869065 /17872

	len =	1112	nex =	3	
	Term	36192	35871	-	0
60	Intr	36608	36279	-	0

1265

	Init	36982	36828	-	0
	>3869065 /28066				
5	len =	202	nex =	1	
	Sngl	58498	58699	+	0
	>3869065 /29368				
10	len =	2470	nex =	7	
	Init	65929	66029	+	0
	Intr	66279	66586	+	0
15	Intr	66669	66764	+	0
	Intr	66861	66938	+	0
	Intr	67036	67212	+	0
	Intr	67884	67952	+	0
	Term	68051	68397	+	0
20	>3869066 /40790				
	len =	2119	nex =	7	
25	Init	35935	36190	+	0
	Intr	36417	36492	+	0
	Intr	36599	36672	+	0
	Intr	36770	36975	+	0
	Intr	37095	37410	+	0
30	Intr	37502	37655	+	0
	Term	37756	38053	+	0
	>3869066 /118207				
35	len =	685	nex =	1	
	Sngl	47112	46433	-	0
	>3869067 /33343				
40	len =	2470	nex =	8	
	Init	22709	22777	+	0
	Intr	22866	22916	+	0
45	Intr	23160	23271	+	0
	Intr	23380	23424	+	0
	Intr	23921	24046	+	0
	Intr	24373	24479	+	0
	Intr	24569	24640	+	0
50	Term	24751	25178	+	0
	>3869067 /20436				
	len =	2137	nex =	5	
55	Init	25230	25389	+	0
	Intr	25682	25841	+	0
	Intr	25964	26519	+	0
	Intr	26612	26883	+	0
60	Term	27050	27366	+	0

2025 RELEASE UNDER E.O. 14176

1266

>3869067 /36010

5	len =	2118	nex =	6	
	Init	29168	29584	+	0
	Intr	29688	30014	+	0
	Intr	30180	30332	+	0
	Intr	30470	30594	+	0
10	Intr	30707	30802	+	0
	Term	30894	31285	+	0

>3869067 /5542

15	len =	1210	nex =	5	
	Term	31596	31244	-	0
	Intr	31767	31669	-	0
	Intr	31964	31870	-	0
20	Intr	32299	32185	-	0
	Init	32449	32382	-	0

>3869067 /12033

25	len =	1338	nex =	3	
	Init	32650	32895	+	0
	Intr	32975	33207	+	0
30	Term	33308	33974	+	0

>3869067 /3488

	len =	1018	nex =	4	
35	Term	34260	34052	-	0
	Intr	34424	34339	-	0
	Intr	34606	34490	-	0
	Init	35069	34689	-	0

40 >3869067 /40511

	len =	2721	nex =	11	
	Init	37516	37614	+	0
45	Intr	37720	37836	+	0
	Intr	37991	38076	+	0
	Intr	38371	38459	+	0
	Intr	38556	38611	+	0
	Intr	38753	38833	+	0
50	Intr	39082	39147	+	0
	Intr	39226	39318	+	0
	Intr	39654	39718	+	0
	Intr	39813	39948	+	0
55	Term	40028	40236	+	0

>3869067 /41451

	len =	933	nex =	1	
60	Sngl	4200	3268	-	0

2025 RELEASE UNDER E.O. 14176

1267

>3869067 /123228

5	len =	999	nex =	3	
	Term	48536	48214	-	0
	Intr	48756	48626	-	0
	Init	49212	48805	-	0

10 >3869067 /25200

	len =	2567	nex =	12	
15	Term	53194	52902	-	0
	Intr	53329	53276	-	0
	Intr	53480	53412	-	0
	Intr	53617	53549	-	0
	Intr	53823	53701	-	0
	Intr	54011	53910	-	0
20	Intr	54167	54093	-	0
	Intr	54334	54244	-	0
	Intr	54762	54425	-	0
	Intr	54939	54853	-	0
	Intr	55148	55058	-	0
25	Init	55468	55243	-	0

>3869067 /17375

	len =	576	nex =	2	
30	Init	69414	69652	+	0
	Term	69732	69989	+	0

35 >3869068 /17883

	len =	535	nex =	1	
	Sngl	16765	16231	-	0

40 >3869068 /30700

	len =	310	nex =	1	
45	Sngl	18011	17704	-	0

>3869068 /29310

	len =	2516	nex =	5	
50	Term	18580	17714	-	0
	Intr	19092	18662	-	0
	Intr	19286	19183	-	0
	Intr	19439	19371	-	0
	Init	20229	20088	-	0

55 >3869068 /112999

	len =	658	nex =	1	
60	Sngl	27286	27943	+	0

2025 RELEASE UNDER E.O. 14176

1268

>3869068 /10694

5	len =	556	nex =	1	
	Sngl	82911	82356	-	0

>3869069 /106959

10	len =	404	nex =	1	
	Sngl	2091	1688	-	0

>3869069 /12707

15	len =	3200	nex =	7	
	Init	30363	30606	+	0
	Intr	31683	31882	+	0
20	Intr	32143	32357	+	0
	Intr	32433	32523	+	0
	Intr	32605	32908	+	0
	Intr	33006	33066	+	0
	Term	33143	33562	+	0

>3869069 /101843

	len =	2862	nex =	10	
30	Init	33963	34200	+	0
	Intr	34320	34353	+	0
	Intr	34515	34655	+	0
	Intr	34751	34838	+	0
	Intr	34915	34995	+	0
35	Intr	35079	35143	+	0
	Intr	35303	35443	+	0
	Intr	35530	35643	+	0
	Intr	35847	35877	+	0
40	Term	36333	36824	+	0

>3869069 /15577

	len =	1904	nex =	3	
45	Term	2612	1677	-	0
	Intr	3118	2808	-	0
	Init	3580	3284	-	0

>3869069 /36656

50	len =	1892	nex =	3	
	Term	2612	1692	-	0
	Intr	3118	2808	-	0
55	Init	3583	3284	-	0

>3869069 /206508

60	len =	1588	nex =	2	
----	-------	------	-------	---	--

0001010203040506070809101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899

1269

Init	39936	40013	+	0
Term	40104	40647	+	0

>3869069 /20484

5	len =	1176	nex =	5	
	Term	52941	52689	-	0
	Intr	53185	53027	-	0
10	Intr	53373	53284	-	0
	Intr	53608	53463	-	0
	Init	53864	53704	-	0

>3869069 /1490

15	len =	1592	nex =	5	
	Term	54883	54300	-	0
	Intr	55072	54965	-	0
20	Intr	55295	55169	-	0
	Intr	55587	55538	-	0
	Init	55891	55681	-	0

>3869069 /5507

25	len =	2151	nex =	9	
	Init	7968	8130	+	0
	Intr	8391	8430	+	0
30	Intr	8506	8598	+	0
	Intr	8713	8761	+	0
	Intr	8841	8947	+	0
	Intr	9024	9111	+	0
	Intr	9197	9400	+	0
35	Intr	9480	9533	+	0
	Term	9618	10118	+	0

>3869069 /33741

40	len =	431	nex =	1	
	Sngl	9688	10112	+	0

>3869070 /12673

45	len =	610	nex =	1	
	Sngl	2159	1550	-	0

>3869071 /486

50	len =	2937	nex =	9	
	Term	11670	11336	-	0
55	Intr	12043	11759	-	0
	Intr	12404	12123	-	0
	Intr	12784	12477	-	0
	Intr	13308	12859	-	0
	Intr	13539	13389	-	0
60	Intr	13791	13627	-	0

2025 RELEASE UNDER E.O. 14176

1270

	Intr	13989	13873	-	0
	Init	14272	14165	-	0
	>3869071	/38585			
5	len =	1270	nex =	5	
	Term	16190	16024	-	0
	Intr	16555	16484	-	0
10	Intr	16728	16660	-	0
	Intr	16951	16821	-	0
	Init	17290	17177	-	0
	>3869071	/9813			
15	len =	1278	nex =	6	
	Term	16190	16019	-	0
	Intr	16395	16323	-	0
20	Intr	16555	16484	-	0
	Intr	16728	16660	-	0
	Intr	16951	16821	-	0
	Init	17296	17177	-	0
	>3869071	/7192			
	len =	839	nex =	1	
	Sngl	30021	29183	-	0
30	>3869071	/22955			
	len =	730	nex =	1	
35	Sngl	38598	37870	-	0
	>3869071	/38211			
	len =	2531	nex =	7	
40	Term	57057	56592	-	0
	Intr	57299	57147	-	0
	Intr	57562	57440	-	0
	Intr	57858	57674	-	0
45	Intr	58188	58099	-	0
	Intr	58480	58304	-	0
	Init	58813	58761	-	0
	>3869072	/94805			
50	len =	862	nex =	3	
	Init	18597	18828	+	0
	Intr	18912	19014	+	0
55	Term	19090	19458	+	0
	>3869072	/17342			
	len =	956	nex =	2	
60					

0369322.261600

1271

	Term	31251	30971	-	0
	Init	31926	31676	-	0
	>3869072 /158846				
5	len =	977	nex =	2	
	Term	31251	30955	-	0
	Init	31911	31676	-	0
10	>3869072 /117248				
	len =	1065	nex =	2	
15	Term	31251	30869	-	0
	Init	31933	31676	-	0
	>3869072 /8695				
20	len =	1011	nex =	2	
	Term	31251	30923	-	0
	Init	31933	31676	-	0
25	>3869072 /20422				
	len =	821	nex =	2	
	Term	31251	31113	-	0
	Init	31933	31676	-	0
30	>3869072 /7107				
	len =	2532	nex =	7	
35	Term	32844	32601	-	0
	Intr	33363	33255	-	0
	Intr	33767	33623	-	0
	Intr	34032	33877	-	0
40	Intr	34209	34148	-	0
	Intr	34421	34295	-	0
	Init	34651	34556	-	0
	>3869072 /18208				
45	len =	861	nex =	1	
	Sngl	44310	44409	+	0
50	>3869072 /26609				
	len =	190	nex =	1	
	Sngl	53681	53499	-	0
55	>3869072 /39314				
	len =	1182	nex =	3	
60	Term	53837	53492	-	0

1272

Intr	54380	54210	-	0
Init	54673	54468	-	0

>3869073 /5685

5	len =	537	nex =	1	
	Sngl	10605	11141	+	0

10 >3869073 /36279

	len =	458	nex =	1	
	Sngl	18587	19044	+	0

15 >3869074 /102795

	len =	632	nex =	1	
--	-------	-----	-------	---	--

20 Sngl 10415 9784 - 0

>3869074 /7236

25 len = 1613 nex = 7

Init 40937 41184 + 0

Intr 41262 41385 + 0

Intr 41618 41726 + 0

Intr 41836 41895 + 0

30 Intr 41988 42053 + 0

Intr 42160 42251 + 0

Term 42351 42549 + 0

>3869074 /7412

35 len = 2576 nex = 11

Term 2325 1773 - 0

Intr 2504 2410 - 0

40 Intr 2685 2589 - 0

Intr 2944 2776 - 0

Intr 3117 3057 - 0

Intr 3305 3204 - 0

Intr 3550 3439 - 0

45 Intr 3788 3634 - 0

Intr 3973 3903 - 0

Intr 4165 4054 - 0

Init 4344 4288 - 0

50 >3869074 /19991

len = 1171 nex = 4

Term 42980 42681 - 0

55 Intr 43303 43190 - 0

Intr 43699 43568 - 0

Init 43851 43806 - 0

>3869074 /16926

60

1273

len = 1210 nex = 4

	Term	42980	42679	-	0
	Intr	43303	43190	-	0
5	Intr	43699	43568	-	0
	Init	43876	43806	-	0

>3869074 /8940

10	len =	3654	nex =	13	
	Term	2325	1767	-	0
	Intr	2504	2410	-	0
	Intr	2685	2589	-	0
15	Intr	2944	2776	-	0
	Intr	3117	3057	-	0
	Intr	3305	3204	-	0
	Intr	3550	3439	-	0
	Intr	3788	3634	-	0
20	Intr	3973	3903	-	0
	Intr	4165	4054	-	0
	Intr	4428	4288	-	0
	Intr	4653	4519	-	0
	Init	5420	4898	-	0

>3869074 /149198

len = 278 nex = 1

30 Sngl 58874 59151 + 0

>3869074 /40315

len = 414 nex = 1

35 Sngl 58874 59287 + 0

>3869074 /154419

40 len = 433 nex = 1

Sngl 58874 59306 + 0

>3869074 /29762

45 len = 854 nex = 1

Sngl 58874 59727 + 0

50 >3869074 /9669

len = 790 nex = 1

55 Sngl 58941 59724 + 0

>3869074 /11244

len = 430 nex = 1

60 Sngl 59119 59548 + 0

2025 RELEASE UNDER E.O. 14176

1275

Intr	12644	12437	-	0
Intr	13278	12727	-	0
Intr	13478	13381	-	0
Init	13959	13755	-	0

5

>3869075 /31648

```
len = 1810    nex = 3
```

10	Init	18786	18937	+	0
	Intr	19201	19479	+	0
	Term	20326	20591	+	0

>3869075 /20274

```

15      len =      1795      nex =          2

```

Init	29646	29969	+	0
Term	30976	31440	+	0

20
>3869075 /9242

```
len = 938    nex = 2
```

25	Init	31907	32136	+	0
	Term	32507	32844	+	0

>3869075 /25655

```
30      len =      896      nex =      2
```

Init	34505	34737	+	0
Term	35146	35400	+	0

35 >3869075 /26281

```
len =      894    nex =      2
```

	Init	34509	34737	+	0
40	Term	35146	35402	+	0

>3869075 /3657

```
len =      490    nex =      1
```

Snq1	36909	37394	+	0
------	-------	-------	---	---

>3869075 /30416

```
50      len =      2410      nex =          7
```

Term	62622	62304	-	0
Intr	62778	62704	-	0
Intr	63019	62876	-	0
Intr	63302	63104	-	0
Intr	63790	63564	-	0
Intr	63930	63882	-	0
Init	64704	64524	-	0

60 >3869075 /1728

1276

```

len = 1844 nex = 5
5 Init 71156 71325 + 0
  Intr 71732 71877 + 0
  Intr 71985 72203 + 0
  Intr 72292 72504 + 0
  Term 72590 72999 + 0

10 >3869075 /42640
    len = 1307 nex = 4
    Init 71693 71877 + 0
    Intr 71985 72203 + 0
    Intr 72292 72504 + 0
    Term 72590 72999 + 0

20 >3869075 /33687
    len = 1549 nex = 2
    Init 7633 8264 + 0
    Term 8616 8707 + 0

25 >3873174 /148790
    len = 1397 nex = 5
    Term 27243 26811 - 0
    Intr 27470 27381 - 0
    Intr 27601 27558 - 0
    Intr 27797 27701 - 0
    Init 28207 27989 - 0

35 >3873174 /101691
    len = 670 nex = 3
    Term 27601 27558 - 0
    Intr 27797 27701 - 0
    Init 28227 27989 - 0

45 >3873174 /38697
    len = 1472 nex = 5
    Term 27243 26768 - 0
    Intr 27470 27381 - 0
    Intr 27601 27558 - 0
    Intr 27797 27701 - 0
    Init 28239 27989 - 0

55 >3873174 /27203
    len = 776 nex = 1
    Sngl 35931 36706 + 0

60 >3873174 /23916

```

1277

	len =	635	nex =	1	
5	Sngl	3852	3218	-	0
	>3873174	/106667			
	len =	572	nex =	1	
10	Sngl	3852	3281	-	0
	>3873174	/7438			
	len =	1037	nex =	1	
15	Sngl	45314	46350	+	0
	>3873174	/2558			
20	len =	1620	nex =	1	
	Sngl	6341	7960	+	0
	>3873174	/40058			
25	len =	2300	nex =	5	
	Term	62162	61975	-	0
	Intr	62557	62385	-	0
30	Intr	63236	63043	-	0
	Intr	63711	63642	-	0
	Init	64274	64013	-	0
	>3873174	/125844			
35	len =	747	nex =	1	
	Sngl	7214	7960	+	0
40	>3873174	/37036			
	len =	1550	nex =	2	
	Term	74393	73816	-	0
45	Init	75365	74610	-	0
	>3873174	/37030			
	len =	2803	nex =	9	
50	Term	97302	97150	-	0
	Intr	97507	97385	-	0
	Intr	97791	97597	-	0
	Intr	98007	97879	-	0
55	Intr	98258	98096	-	0
	Intr	98487	98352	-	0
	Intr	98916	98575	-	0
	Intr	99091	99011	-	0
	Init	99952	99437	-	0
60					

1278

>3885325 /21742

len = 790 nex = 3

5	Term	19497	19031	-	0
	Intr	19671	19581	-	0
	Init	19813	19745	-	0

>3885325 /37160

10 len = 2950 nex = 10

	Term	19497	19026	-	0
	Intr	19671	19581	-	0
15	Intr	19813	19745	-	0
	Intr	20265	20078	-	0
	Intr	20516	20382	-	0
	Intr	20681	20607	-	0
	Intr	21008	20768	-	0
20	Intr	21560	21482	-	0
	Intr	21743	21651	-	0
	Init	21973	21878	-	0

>3885325 /8342

25 len = 2611 nex = 10

	Term	28061	27841	-	0
	Intr	28235	28145	-	0
30	Intr	28405	28340	-	0
	Intr	28839	28793	-	0
	Intr	29013	28929	-	0
	Intr	29188	29102	-	0
	Intr	29519	29446	-	0
35	Intr	29749	29666	-	0
	Intr	30098	30059	-	0
	Init	30451	30230	-	0

>3885325 /14041

40 len = 1182 nex = 2

	Init	31479	31619	+	0
	Term	32162	32660	+	0

45

>3885325 /29321

len = 936 nex = 2

50	Init	31484	31619	+	0
	Term	32162	32419	+	0

>3885325 /23675

55 len = 1196 nex = 2

	Init	31488	31619	+	0
	Term	32162	32683	+	0

60 >3885325 /7008

20250802 09:22:00

1279

	len =	996	nex =	2	
	Init	31541	31619	+	0
5	Term	32162	32536	+	0
	>3885325 /35542				
	len =	4002	nex =	3	
10	Term	76796	76680	-	0
	Intr	77285	77136	-	0
	Init	77644	77555	-	0
	15 >3885325 /8450				
	len =	1248	nex =	3	
	Init	94767	95217	+	0
20	Intr	95576	95626	+	0
	Term	95755	96014	+	0
	>3885325 /30446				
	len =	942	nex =	2	
25	Term	98533	97856	-	0
	Init	98797	98628	-	0
	30 >3892698 /5111				
	len =	852	nex =	1	
	Sngl	49208	48945	-	0
35	>3892698 /35338				
	len =	2170	nex =	2	
40	Term	47914	47683	-	0
	Init	49208	49126	-	0
	>3892698 /36849				
45	len =	772	nex =	1	
	Sngl	51041	50270	-	0
	>3892698 /18749				
50	len =	1672	nex =	2	
	Term	56961	55850	-	0
	Init	57521	57295	-	0
55	>3892698 /11009				
	len =	3236	nex =	6	
60	Term	54492	54297	-	0

2025-02-25 09:00:00

1280

	Intr	54670	54612	-	0
	Intr	54925	54776	-	0
	Intr	55085	54993	-	0
	Intr	56961	56851	-	0
5	Init	57532	57295	-	0
	>3892698	/5348			
	len =	535	nex =	1	
10	Sngl	74771	74237	-	0
	>3894156	/11351			
	len =	2097	nex =	4	
	Init	37638	37696	+	0
	Intr	37801	38037	+	0
	Intr	38134	38351	+	0
20	Term	38433	38887	+	0
	>3894156	/18246			
	len =	2050	nex =	9	
	Init	44721	44808	+	0
	Intr	44904	44960	+	0
	Intr	45125	45231	+	0
	Intr	45327	45398	+	0
30	Intr	45557	45610	+	0
	Intr	45697	45772	+	0
	Intr	45919	45980	+	0
	Intr	46137	46196	+	0
	Term	46455	46618	+	0
35	>3894156	/107424			
	len =	2332	nex =	9	
	Init	44688	44808	+	0
	Intr	44904	44960	+	0
	Intr	45125	45231	+	0
	Intr	45327	45398	+	0
	Intr	45557	45610	+	0
45	Intr	45697	45772	+	0
	Intr	45919	45980	+	0
	Intr	46137	46196	+	0
	Term	46455	47019	+	0
50	>3894156	/114042			
	len =	1556	nex =	2	
	Term	47511	46634	-	0
55	Init	48189	47592	-	0
	>3894156	/38635			
	len =	2276	nex =	5	
60					

1281

	Init	68187	68540	+	0
	Intr	68715	68747	+	0
	Intr	68929	69157	+	0
	Intr	69743	69869	+	0
5	Term	69944	70462	+	0
	>3894156 /41104				
10	len =	1762	nex =	5	
	Init	68444	68540	+	0
	Intr	68715	68747	+	0
	Intr	68929	69157	+	0
	Intr	69743	69869	+	0
15	Term	69944	70205	+	0
	>3894179 /36320				
20	len =	158	nex =	1	
	Sngl	19935	19778	-	0
	>3894179 /13608				
25	len =	1419	nex =	3	
	Init	3877	4048	+	0
	Intr	4194	4398	+	0
	Term	5037	5295	+	0
30	>3894179 /11854				
	len =	2214	nex =	6	
35	Term	38528	38322	-	0
	Intr	38784	38647	-	0
	Intr	39404	39281	-	0
	Intr	39652	39529	-	0
	Intr	39874	39742	-	0
40	Init	40535	40056	-	0
	>3894179 /125677				
45	len =	676	nex =	2	
	Init	6794	6915	+	0
	Term	7018	7469	+	0
	>3894179 /9236				
50	len =	575	nex =	1	
	Sngl	76198	76772	+	0
55	>3894179 /15522				
	len =	527	nex =	1	
60	Sngl	76584	76058	-	0

1282

>3894179 /31116

```
len = 459    nex = 1
```

5 Snq1 80390 80848 + 0

>3894179 /30800

```
len =      409      nex =      1
```

10					
	Snq1	83631	83223	-	0

>3894179 /37949

```
15      len =      1100      nex =      4
```

Term	83667	83218	-	0
------	-------	-------	---	---

Intr	83835	83752	-	0
------	-------	-------	---	---

Intr	83992	83914	-	0
------	-------	-------	---	---

20	Init	84302	84098	-	0
----	------	-------	-------	---	---

>3894179 /11537

```
len =      953    nex =      1
```

25					
	Snq1	8992	9944	+	0

>3927822 /32642

```
30      len =      500      nex =      1
```

Snq1	1179	680	-	0
------	------	-----	---	---

>3927822 /2058

```

35      len =      927      nex =      1

```

Snq1	26615	25689	-	0
------	-------	-------	---	---

40 >3927822 /35476

```
len = 589    nex = 2
```

Term	28078	27741	-	0
------	-------	-------	---	---

45	Init	28329	28179	-	0
----	------	-------	-------	---	---

>3927822 /31457

```
len = 2687    nex = 10
```

50

Term	28078	27729	-	0
------	-------	-------	---	---

Intr	28328	28179	-	0
------	-------	-------	---	---

Intr	28740	28637	-	0
------	-------	-------	---	---

Intr	28940	28833	-	0
------	-------	-------	---	---

55	Intr	29186	29082	-	0
----	------	-------	-------	---	---

Intr	29361	29288	-	0
------	-------	-------	---	---

Intr	29506	29446	-	0
------	-------	-------	---	---

Intr	29658	29584	-	0
------	-------	-------	---	---

Intr	29879	29769	-	0
------	-------	-------	---	---

60	Init	30415	29968	-	0
----	------	-------	-------	---	---

1283

>3927822 /29679

5	len =	2351	nex =	8	
	Init	49567	49905	+	0
	Intr	49994	50053	+	0
	Intr	50131	50202	+	0
	Intr	50276	50397	+	0
10	Intr	50479	50619	+	0
	Intr	50710	50817	+	0
	Intr	50893	51145	+	0
	Term	51224	51814	+	0

15 >3927822 /22382

```
len =      660    nex =      1
Snq1  66099    65440      -      0
```

>3927822 /10159

	len =	1250	nex =	6	
25	Term	85753	85484	-	0
	Intr	85900	85833	-	0
	Intr	86011	85978	-	0
	Intr	86139	86108	-	0
	Intr	86442	86391	-	0
30	Init	86733	86530	-	0

>3927822 /948

```

35      len =      1606      nex =          1
      Sngl 88569  86964      -          0

```

>3928074 /7805

40	len =	2278	nex =	5	
	Init	10584	10991	+	0
	Intr	11688	11817	+	0
	Intr	12087	12157	+	0
45	Intr	12246	12326	+	0
	Term	12682	12861	+	0

>3928074 /17830

```

50      len =      819      nex =      2
        Init 15343 15421      +      0
        Term 15807 16161      +      0

```

55 >3928074 /38129

	len =	1666	nex =	4	
60	Term	5915	5673	-	0
	Intr	6499	6134	-	0

1284

Intr	7048	6578	-	0
Init	7338	7135	-	0

>3928074 /39558

5	len =	2141	nex =	3	
	Init	75510	76133	+	0
	Intr	76482	76604	+	0
10	Term	76976	77204	+	0

>3928074 /20810

15	len =	1514	nex =	5	
	Term	79125	78844	-	0
	Intr	79336	79217	-	0
	Intr	79585	79475	-	0
	Intr	79927	79717	-	0
20	Init	80038	80002	-	0

>3928074 /11662

25	len =	898	nex =	2	
	Term	8846	8488	-	0
	Init	9385	9139	-	0

>3980374 /34873

30	len =	738	nex =	2	
	Term	116215	115949	-	0
	Init	116686	116328	-	0

>3980374 /17240

35	len =	1948	nex =	3	
40	Init	13701	14199	+	0
	Intr	14385	14654	+	0
	Term	14795	15648	+	0

>3980374 /13720

45	len =	1219	nex =	3	
	Term	16100	15850	-	0
	Intr	16298	16222	-	0
50	Init	17068	16911	-	0

>3980374 /5112

55	len =	1528	nex =	5	
	Init	17481	17533	+	0
	Intr	17782	17834	+	0
	Intr	18312	18350	+	0
	Intr	18477	18546	+	0
60	Term	18662	18881	+	0

>3980374 /25736

5	len =	1510	nex =	5	
	Init	17406	17533	+	0
	Intr	17782	17834	+	0
	Intr	18312	18350	+	0
	Intr	18477	18546	+	0
10	Term	18662	18914	+	0

>3980374 /25828

15	len =	691	nex =	1	
	Sngl	25933	26623	+	0

>3980374 /38727

20	len =	950	nex =	2	
	Init	27265	27698	+	0
	Term	27792	28214	+	0

25 >3980374 /6528

	len =	1035	nex =	2	
	Init	34104	34489	+	0
30	Term	34608	35138	+	0

>3980374 /24361

35	len =	898	nex =	2	
	Init	40839	41193	+	0
	Term	41271	41736	+	0

>3980374 /39038

40	len =	981	nex =	3	
	Init	4299	4497	+	0
	Intr	4750	4890	+	0
45	Term	5212	5279	+	0

>3980374 /6157

50	len =	741	nex =	2	
	Init	45228	45692	+	0
	Term	45776	45968	+	0

>3980374 /23439

55	len =	2314	nex =	7	
	Init	46866	46968	+	0
	Intr	47598	47735	+	0
60	Intr	47926	48071	+	0

1286

Intr	48238	48357	+	0
Intr	48453	48665	+	0
Intr	48761	48841	+	0
Term	48926	49179	+	0

5

>3980374 /33426

len = 1249 nex = 4

10	Init	47926	48071	+	0
	Intr	48238	48357	+	0
	Intr	48453	48665	+	0
	Term	48761	48842	+	0

15 >3980374 /2618

len = 1394 nex = 5

20	Term	54808	54506	-	0
	Intr	55007	54901	-	0
	Intr	55303	55087	-	0
	Intr	55595	55388	-	0
	Init	55899	55798	-	0

25 >3980374 /14555

len = 1516 nex = 5

30	Term	57704	57309	-	0
	Intr	57909	57803	-	0
	Intr	58204	57988	-	0
	Intr	58508	58301	-	0
	Init	58824	58614	-	0

35 >3980374 /1637

len = 1285 nex = 5

40	Term	66552	66306	-	0
	Intr	66749	66643	-	0
	Intr	67079	66860	-	0
	Intr	67374	67167	-	0
	Init	67590	67476	-	0

45 >3980374 /342

len = 1796 nex = 4

50	Init	8035	8458	+	0
	Intr	8947	9099	+	0
	Intr	9183	9280	+	0
	Term	9375	9830	+	0

>3980374 /28554

55 len = 1450 nex = 4

60	Init	8140	8458	+	0
	Intr	8947	9099	+	0
	Intr	9183	9280	+	0

2025 RELEASE UNDER E.O. 14176

1287

	Term	9375	9586	+	0
	>3980374	/21655			
5	len =	310	nex =	1	
	Sngl	9418	9723	+	0
	>3983533	/40760			
10	len =	811	nex =	2	
	Init	55705	55965	+	0
	Term	56057	56515	+	0
15	>3985931	/11375			
	len =	1235	nex =	4	
20	Term	27019	26645	-	0
	Intr	27179	27114	-	0
	Intr	27323	27258	-	0
	Init	27879	27598	-	0
25	>3985931	/32856			
	len =	1429	nex =	1	
	Sngl	40537	39109	-	0
30	>3985931	/4623			
	len =	2005	nex =	9	
35	Term	43824	43509	-	0
	Intr	43989	43913	-	0
	Intr	44159	44065	-	0
	Intr	44295	44251	-	0
	Intr	44544	44399	-	0
40	Intr	44741	44654	-	0
	Intr	44873	44843	-	0
	Intr	45006	44981	-	0
	Init	45513	45090	-	0
45	>3985931	/109141			
	len =	670	nex =	1	
	Sngl	73609	72960	-	0
50	>3985931	/19506			
	len =	692	nex =	1	
55	Sngl	9471	10162	+	0
	>3985931	/119485			
	len =	629	nex =	1	
60					

1288

	Sngl	9473	10101	+	0
	>3985931	/4831			
5	len =	565	nex =	1	
	Sngl	9591	10155	+	0
	>3985933	/39005			
10	len =	4007	nex =	11	
	Init	28319	28431	+	0
	Intr	28805	28887	+	0
15	Intr	28996	29125	+	0
	Intr	29224	29303	+	0
	Intr	29443	29529	+	0
	Intr	29623	29747	+	0
	Intr	29839	29899	+	0
20	Intr	30132	30209	+	0
	Intr	30301	30406	+	0
	Intr	30490	30635	+	0
	Term	30833	31065	+	0
25	>3985934	/26805			
	len =	677	nex =	1	
	Sngl	19613	18937	-	0
30	>3985934	/25275			
	len =	2482	nex =	9	
35	Init	37180	37552	+	0
	Intr	37890	38045	+	0
	Intr	38268	38336	+	0
	Intr	38434	38604	+	0
	Intr	38687	38830	+	0
40	Intr	38925	39083	+	0
	Intr	39196	39261	+	0
	Intr	39362	39421	+	0
	Term	39502	39661	+	0
45	>3985934	/91760			
	len =	1371	nex =	6	
	Init	38434	38604	+	0
50	Intr	38687	38830	+	0
	Intr	38925	39083	+	0
	Intr	39196	39261	+	0
	Intr	39362	39421	+	0
	Term	39502	39658	+	0
55	>3985934	/39478			
	len =	3370	nex =	4	
60	Init	3984	4155	+	0

0505326-061600

1289

Intr	4879	5952	+	0
Intr	6176	6354	+	0
Term	6691	7351	+	0

5 >3985934 /31667

len = 1184 nex = 2

Init	40094	40516	+	0
Term	40861	41277	+	0

>3985934 /118260

len = 926 nex = 2

Init	40294	40516	+	0
Term	40861	41219	+	0

>3985934 /13962

len = 568 nex = 1

Sngl 42221 41654 - 0

25 >3985934 /32925

len = 447 nex = 1

Sngl 43700 43254 - 0

30 >3985934 /14816

len = 1140 nex = 4

Init	48454	48662	+	0
Intr	48749	48979	+	0
Intr	49063	49263	+	0
Term	49374	49593	+	0

40 >3985949 /158734

len = 624 nex = 1

Sngl 10055 10678 + 0

45 >3985949 /15880

len = 1150 nex = 2

Term	19933	19551	-	0
Init	20691	20400	-	0

>3985949 /17909

55 len = 3262 nex = 8

Init	51683	52348	+	0
Intr	52522	53058	+	0
Intr	53154	53240	+	0
Intr	53329	53399	+	0

60

1290

Intr	53470	53546	+	0
Intr	53652	53799	+	0
Intr	53897	54058	+	0
Term	54365	54944	+	0

5

>3985949 /26535

```
len = 1492    nex = 5
```

10	Init	62051	62540	+	0
	Intr	62742	62879	+	0
	Intr	62957	63043	+	0
	Intr	63118	63185	+	0
	Term	63285	63542	+	0

15

>3985950 /102285

```
len =      436      nex =      1
```

20	Sngl	22343	22778	+	0
----	------	-------	-------	---	---

>3985952 /9393

```
len = 3254    nex = 6
```

25	Init	12815	13266	+	0
	Intr	13492	13653	+	0
	Intr	13771	13957	+	0
	Intr	14238	14663	+	0
30	Intr	15037	15444	+	0
	Term	15529	16068	+	0

>3985952 /11984

```
35      len =      2739      nex =      12
```

Init	20557	20824	+	0
Intr	20898	20956	+	0
Intr	21064	21139	+	0
Intr	21217	21288	+	0
Intr	21367	21454	+	0
Intr	21528	21780	+	0
Intr	21897	22012	+	0
Intr	22111	22331	+	0
Intr	22422	22453	+	0
Intr	22546	22648	+	0
Intr	22763	22943	+	0
Term	23026	23295	+	0

50 >3985952 /207193

```
len = 204    nex = 1
```

Snq1 23109 23312 + 0

55

>3985952 /3151

```
len =      537    nex =      2
```

60 Term 40208 40095 - 0

1291

Init	40631	40491	-	0
------	-------	-------	---	---

>3985952 /113501

5	len =	910	nex =	2
---	-------	-----	-------	---

Term	40208	39985	-	0
Init	40888	40491	-	0

10 >3985952 /38966

len =	1279	nex =	2
-------	------	-------	---

15	Term	40208	40095	-	0
	Init	41373	40491	-	0

>3985952 /35731

len =	2127	nex =	10
-------	------	-------	----

20	Term	41847	41632	-	0
	Intr	41987	41928	-	0
	Intr	42146	42067	-	0
	Intr	42309	42227	-	0
25	Intr	42525	42409	-	0
	Intr	42673	42596	-	0
	Intr	42917	42848	-	0
	Intr	43184	43072	-	0
	Intr	43373	43272	-	0
30	Init	43758	43458	-	0

>3985952 /2064

len =	2367	nex =	5
-------	------	-------	---

35	Init	49630	49929	+	0
	Intr	50622	50752	+	0
	Intr	51041	51168	+	0
	Intr	51324	51530	+	0
40	Term	51678	51996	+	0

>3985952 /41430

len =	2159	nex =	6
-------	------	-------	---

45	Init	52202	52485	+	0
	Intr	52919	53097	+	0
	Intr	53191	53274	+	0
	Intr	53631	53700	+	0
50	Intr	53796	53929	+	0
	Term	54020	54360	+	0

>3985952 /7925

55	len =	1873	nex =	1
----	-------	------	-------	---

Sngl	61124	62996	+	0
------	-------	-------	---	---

>3985952 /12726

60

1292

	len =	291	nex =	1	
	Sngl	62743	63033	+	0
5	>3985952	/6129			
	len =	629	nex =	1	
10	Sngl	70947	71575	+	0
	>3985954	/850			
	len =	2781	nex =	9	
15	Term	22960	22894	-	0
	Intr	23249	23031	-	0
	Intr	23819	23641	-	0
	Intr	24255	24075	-	0
	Intr	24624	24497	-	0
20	Intr	24745	24704	-	0
	Intr	24912	24833	-	0
	Intr	25199	25105	-	0
	Init	25674	25295	-	0
25	>3985954	/39002			
	len =	2234	nex =	3	
	Init	51257	51740	+	0
30	Intr	51829	52098	+	0
	Term	52600	53490	+	0
	>3985955	/33506			
35	len =	2022	nex =	6	
	Term	12286	12055	-	0
	Intr	12479	12381	-	0
	Intr	12746	12565	-	0
40	Intr	12936	12836	-	0
	Intr	13257	13097	-	0
	Init	14076	13714	-	0
	>3985955	/151935			
45	len =	175	nex =	1	
	Sngl	55158	55332	+	0
50	>3985955	/22697			
	len =	1494	nex =	6	
	Term	56280	55912	-	0
55	Intr	56493	56416	-	0
	Intr	56671	56580	-	0
	Intr	56947	56897	-	0
	Intr	57238	57184	-	0
60	Init	57405	57326	-	0

1293

>3985955 /42952

```
len = 1548    nex = 4
```

5	Init	61	157	+	0
	Intr	630	706	+	0
	Intr	796	868	+	0
	Term	974	1608	+	0

10 >3985955 /33522

```
len = 1783    nex = 0
```

>3985957 /108603

```
len =      472    nex =      1
```

Sngl	14977	15448	+	0
------	-------	-------	---	---

20 >3985957 /38105

```
len = 2110    nex = 6
```

25	Init	2224	2751	+	0
	Intr	3237	3409	+	0
	Intr	3511	3559	+	0
	Intr	3661	3827	+	0
	Intr	3920	4003	+	0
	Term	4106	4329	+	0

30 >3985957 /99783

```
len = 770    nex = 1
```

35 Snq1 32397 33166 + 0

>3985957 /35856

```
len = 2121 nex = 6
```

	Init	51112	51715	+	0
	Intr	52102	52204	+	0
	Intr	52327	52415	+	0
	Intr	52619	52710	+	0
45	Intr	52792	52889	+	0
	Term	52976	53232	+	0

>3985958 / 36809

```
50      len =      1903      nex =          4
```

	Init	32457	32776	+	0
	Intr	32856	32932	+	0
	Intr	33601	33919	+	0
55	Term	33996	34359	+	0

>3985958 /36602

```

len = 2296  nex = 5
60

```

1294

	Init	5089	5425	+	0
	Intr	6229	6296	+	0
	Intr	6383	6514	+	0
	Intr	6654	6769	+	0
5	Term	6876	7384	+	0

>3985958 /25325

10	len =	1821	nex =	6	
	Init	52395	52503	+	0
	Intr	53039	53117	+	0
	Intr	53241	53312	+	0
	Intr	53437	53619	+	0
15	Intr	53708	53783	+	0
	Term	53955	54215	+	0

>3985958 /30108

20	len =	983	nex =	4	
	Term	61360	61218	-	0
	Intr	61713	61665	-	0
	Intr	61976	61828	-	0
25	Init	62200	62078	-	0

>3985958 /11295

30	len =	1274	nex =	4	
	Term	73348	72789	-	0
	Intr	73499	73432	-	0
	Intr	73829	73580	-	0
	Init	74062	73922	-	0

>3985958 /17636

	len =	585	nex =	1	
40	Sngl	7885	7301	-	0

>3985958 /16784

45	len =	1046	nex =	2	
	Term	80867	80492	-	0
	Init	81076	80939	-	0

>3985958 /97883

50	len =	1074	nex =	4	
	Init	8598	8696	+	0
	Intr	8960	9035	+	0
55	Intr	9202	9513	+	0
	Term	9628	9671	+	0

>4003353 /11111

60	len =	1241	nex =	4	
----	-------	------	-------	---	--

1295

	Term	18714	18255	-	0
	Intr	18983	18813	-	0
	Intr	19249	19172	-	0
5	Init	19495	19337	-	0
	>4003353 /465				
	len =	717	nex =	2	
10	Init	23467	23626	+	0
	Term	23709	24183	+	0
	>4003353 /22440				
15	len =	2230	nex =	3	
	Term	46405	45777	-	0
	Intr	46873	46618	-	0
20	Init	48003	47263	-	0
	>4003353 /12261				
	len =	2365	nex =	3	
25	Term	46405	45643	-	0
	Intr	46873	46618	-	0
	Init	47683	47263	-	0
30	>4003353 /21954				
	len =	2197	nex =	8	
	Term	3077	2689	-	0
35	Intr	3298	3253	-	0
	Intr	3684	3577	-	0
	Intr	3834	3775	-	0
	Intr	4211	4130	-	0
	Intr	4401	4286	-	0
40	Intr	4622	4494	-	0
	Init	4885	4714	-	0
	>4003353 /39309				
45	len =	2590	nex =	8	
	Init	69646	69794	+	0
	Intr	69914	70187	+	0
	Intr	70279	70334	+	0
50	Intr	70442	70590	+	0
	Intr	70894	71049	+	0
	Intr	71189	71234	+	0
	Intr	71571	71637	+	0
	Term	71818	71894	+	0
55	>4003353 /155178				
	len =	452	nex =	1	
60	Sngl	72936	72485	-	0

>4003353 /33702

5 len = 298 nex = 1
 Sngl 73011 72714 - 0

>4003353 /21040

10 len = 511 nex = 1
 Sngl 73016 72506 - 0

>4006815 /11214

15 len = 2790 nex = 10
 Init 107144 107253 + 0
 Intr 107333 107388 + 0
 20 Intr 107462 107594 + 0
 Intr 107672 107746 + 0
 Intr 107835 107950 + 0
 Intr 108022 108073 + 0
 25 Intr 108174 108225 + 0
 Intr 108306 108363 + 0
 Intr 108456 108554 + 0
 Term 108635 108985 + 0

>4006815 /36277

30 len = 2791 nex = 10
 Init 107144 107253 + 0
 Intr 107333 107388 + 0
 35 Intr 107462 107594 + 0
 Intr 107672 107746 + 0
 Intr 107835 107950 + 0
 Intr 108022 108073 + 0
 40 Intr 108174 108225 + 0
 Intr 108306 108363 + 0
 Intr 108456 108554 + 0
 Term 108635 108992 + 0

>4006815 /10529

45 len = 1092 nex = 0

>4006815 /2886

50 len = 624 nex = 1
 Sngl 114766 115389 + 0

>4006815 /36965

55 len = 2350 nex = 9
 Init 13554 13743 + 0
 Intr 13824 13879 + 0
 60 Intr 13954 14112 + 0

1297

	Intr	14196	14295	+	0
	Intr	14381	14523	+	0
	Intr	14714	14819	+	0
	Intr	15089	15227	+	0
5	Intr	15323	15591	+	0
	Term	15672	15895	+	0
	>4006815		/2240		
10	len =	2135	nex =	9	
	Init	18070	18214	+	0
	Intr	18297	18391	+	0
	Intr	18473	18554	+	0
15	Intr	18619	18681	+	0
	Intr	18791	18845	+	0
	Intr	18932	19014	+	0
	Intr	19321	19416	+	0
	Intr	19606	19665	+	0
20	Term	19874	20200	+	0
	>4006815		/18652		
	len =	2128	nex =	9	
25	Init	18079	18214	+	0
	Intr	18297	18391	+	0
	Intr	18473	18554	+	0
	Intr	18619	18715	+	0
30	Intr	18791	18845	+	0
	Intr	18932	19014	+	0
	Intr	19321	19416	+	0
	Intr	19606	19665	+	0
	Term	19874	20202	+	0
35	>4006815		/29981		
	len =	550	nex =	1	
40	Sngl	47608	47066	-	0
	>4006815		/12018		
	len =	594	nex =	1	
45	Sngl	48483	47890	-	0
	>4006815		/33863		
50	len =	207	nex =	1	
	Sngl	53635	53429	-	0
	>4006815		/6541		
55	len =	1765	nex =	2	
	Term	54408	53438	-	0
	Init	55202	54521	-	0
60					

1298

>4006885 /38141

```
len = 1902    nex = 3
```

5	Init	102102	102792	+	0
	Intr	103399	103488	+	0
	Term	103569	103636	+	0

>4006885 /21999

```
10      len =      1291      nex =      2
```

Init	107190	107560	+	0
Term	107858	108480	+	0

15 >4006885 /143475

```
len = 1163    nex = 2
```

20	Term	113830	113699	-	0
	Init	114861	114488	-	0

>4006885 / 36845

```
25      len =      1815      nex =          7
```

Init	127608	127748	+	0
Intr	127830	127937	+	0
Intr	128025	128163	+	0
Intr	128267	128437	+	0
Intr	128516	128756	+	0
Intr	128833	128924	+	0
Term	129014	129422	+	0

35 >4006885 /95065

```
len =      238    nex =      1
```

Snq1	143397	143160	-	0
------	--------	--------	---	---

40 >4006885 /40968

```
len = 1059    nex = 6
```

45	Term	166624	166434	-	0
	Intr	166800	166712	-	0
		166946	166883	-	0
	Intr	167137	167035	-	0
	Intr	167317	167214	-	0
50	Init	167492	167421	-	0

>4006885 /36701

```
len = 1410    nex = 1
```

55

Snq1	26044	24635	-	0
------	-------	-------	---	---

>4006885 /30175

```
60      len =      1665      nex =          5
```


1300

Intr	40036	40369	+	0
Term	40452	40792	+	0

5 >4038029 /12043

```
len = 1878    nex = 11
```

	Term	23390	23308	-	0
	Intr	23553	23476	-	0
10	Intr	23748	23695	-	0
	Intr	23896	23833	-	0
	Intr	24208	24163	-	0
	Intr	24367	24295	-	0
	Intr	24518	24447	-	0
15	Intr	24700	24608	-	0
	Intr	24842	24790	-	0
	Intr	25011	24921	-	0
	Init	25171	25089	-	0

20 >4038029 /12568

```
len = 2394    nex = 11
```

25	Term	23390	23002	-	0
	Intr	23553	23476	-	0
	Intr	23748	23695	-	0
	Intr	23896	23833	-	0
	Intr	24208	24163	-	0
30	Intr	24367	24295	-	0
	Intr	24518	24447	-	0
	Intr	24700	24608	-	0
	Intr	24842	24790	-	0
	Intr	25011	24921	-	0
	Intr	25395	25089	-	0

35 >4038029 /31175

```
len =      719      nex =      2
```

40	Term	41247	40834	-	0
	Init	41552	41441	-	0

>4038029 /205695

```
45      len =      575      nex =      2
```

Term	46021	45667	-	0
Init	46241	46125	-	0

50 >4038029 /364

```
len =      598      nex =      2
```

	Term	46021	45663	-	0
55	Init	46260	46125	-	0

>4038029 /6995

```

60      len =      767      nex =      2

```

1301

Term	57543	57164	-	0
Init	57930	57799	-	0

5	>4038029	/9804		
	len =	2193	nex =	3
	Init	65054	65282	+
	Intr	65599	65995	+
10	Term	66507	66778	+
	>4049332	/31275		
	len =	1090	nex =	1
15	Sngl	36708	37796	+
	>4049332	/12232		
20	len =	1707	nex =	3
	Init	38921	39520	+
	Intr	39710	39871	+
	Term	40057	40627	+
25	>4049332	/31460		
	len =	1016	nex =	2
30	Init	39710	39871	+
	Term	40057	40719	+
	>4049332	/37529		
35	len =	2144	nex =	3
	Term	63854	63126	-
	Intr	64463	63966	-
	Init	65269	65079	-
40	>4049332	/34310		
	len =	1773	nex =	4
45	Term	70001	69714	-
	Intr	70410	70363	-
	Intr	71182	71025	-
	Init	71486	71294	-
50	>4049332	/28012		
	len =	1066	nex =	2
	Init	77171	77322	+
55	Term	77644	78233	+
	>4049332	/16844		
60	len =	957	nex =	3

1302

Term	90208	89783	-	0
Intr	90525	90453	-	0
Init	90723	90633	-	0

5 >4056429 /33126

```
len = 1510    nex = 5
```

10	Init	10968	11580	+	0
	Intr	11710	11780	+	0
	Intr	11934	12053	+	0
	Intr	12141	12259	+	0
	Term	12393	12475	+	0

15 >4056429 /3465

```
len = 1407    nex = 4
```

	Term	18080	17627	-	0
20	Intr	18319	18176	-	0
	Intr	18697	18419	-	0
	Init	19033	18817	-	0

>4056429 /26826

25	len =	1648	nex =	5	
	Term	27443	26816	-	0
	Intr	27669	27526	-	0
30	Intr	27875	27753	-	0
	Intr	28125	27964	-	0
	Init	28463	28218	-	0

>4056429 /92839

```

35      len =      1320      nex =          4
          Term 29933 29790 - 0
          Intr 30139 30074 - 0
          Intr 30433 30254 - 0
40      Init 30662 30509 - 0

```

>4056429 /6245

45	len =	1515	nex =	5	
	Term	33847	33332	-	0
	Intr	34084	33941	-	0
	Intr	34319	34185	-	0
	Intr	34571	34422	-	0
50	Init	34846	34705	-	0

>4056429 /10721

```

55      len =      2590      nex =          4
        Term 47922  47487      -          0
        Intr 48181  48035      -          0
        Intr 49877  49731      -          0
60      Init 50069  49958      -          0

```


1303

>4056429 /36437

len = 1094 nex = 1

5 Sngl 59727 60545 + 0

>4056429 /15450

len = 1313 nex = 2

10 Init 74846 75064 + 0

Term 75185 75669 + 0

>4056429 /18845

15 len = 2571 nex = 9

Init 94527 94779 + 0

Intr 94873 94949 + 0

20 Intr 95041 95163 + 0

Intr 95819 95923 + 0

Intr 96020 96092 + 0

Intr 96192 96263 + 0

25 Intr 96369 96521 + 0

Intr 96649 96772 + 0

Term 96860 97097 + 0

>4056476 /31973

30 len = 570 nex = 1

Sngl 105977 106546 + 0

>4056476 /39206

35 len = 1908 nex = 4

Init 107066 107469 + 0

Intr 107817 107846 + 0

40 Intr 107898 108112 + 0

Term 108463 108973 + 0

>4056476 /205753

45 len = 99 nex = 1

Sngl 107310 107212 - 0

>4056476 /98881

50 len = 1823 nex = 9

Term 9508 9155 - 0

Intr 9685 9625 - 0

55 Intr 9877 9786 - 0

Intr 10054 9975 - 0

Intr 10183 10147 - 0

Intr 10330 10271 - 0

Intr 10462 10417 - 0

60 Intr 10672 10540 - 0

20250326.061603

1304

	Init	10977	10757	-	0
	>4056476 /15831				
5	len =	1631	nex =	4	
	Init	11178	11501	+	0
	Intr	11809	11949	+	0
	Intr	12030	12302	+	0
10	Term	12389	12808	+	0
	>4056476 /11642				
15	len =	1479	nex =	3	
	Term	112566	112276	-	0
	Intr	112710	112642	-	0
	Init	113754	113600	-	0
20	>4056476 /29782				
	len =	1167	nex =	4	
	Term	17151	17095	-	0
25	Intr	17449	17246	-	0
	Intr	17819	17548	-	0
	Init	18261	18164	-	0
	>4056476 /8397				
30	len =	1454	nex =	3	
	Term	17151	16982	-	0
	Intr	17449	17246	-	0
35	Init	17819	17548	-	0
	>4056476 /33020				
40	len =	1690	nex =	6	
	Term	39931	39868	-	0
	Intr	40166	40071	-	0
	Intr	40433	40246	-	0
	Intr	40634	40539	-	0
45	Intr	40814	40775	-	0
	Init	41554	41439	-	0
	>4056476 /113469				
50	len =	2530	nex =	6	
	Term	39931	39376	-	0
	Intr	40166	40071	-	0
	Intr	40433	40246	-	0
55	Intr	40634	40539	-	0
	Intr	40814	40775	-	0
	Init	41552	41439	-	0
	>4056476 /43008				
60					

1305

	len =	2506	nex =	6	
	Term	39931	39406	-	0
	Intr	40166	40071	-	0
5	Intr	40433	40246	-	0
	Intr	40634	40539	-	0
	Intr	40814	40775	-	0
	Init	41552	41439	-	0
10	>4056476		/8849		
	len =	1775	nex =	4	
	Term	4259	4148	-	0
15	Intr	4695	4357	-	0
	Intr	5068	4967	-	0
	Init	5605	5183	-	0
20	>4056476		/36303		
	len =	1786	nex =	4	
	Term	4259	4148	-	0
	Intr	4695	4357	-	0
25	Intr	5068	4967	-	0
	Init	5616	5183	-	0
	>4056476		/9001		
30	len =	1248	nex =	1	
	Sngl	66616	65369	-	0
	>4056476		/16393		
35	len =	2532	nex =	6	
	Term	96093	95777	-	0
	Intr	96329	96178	-	0
40	Intr	96583	96422	-	0
	Intr	97005	96654	-	0
	Intr	97159	97097	-	0
	Init	98308	98024	-	0
45	>4063730		/13660		
	len =	1755	nex =	3	
	Init	60286	60642	+	0
50	Intr	61422	61490	+	0
	Term	61594	62040	+	0
	>4063730		/21240		
55	len =	2170	nex =	5	
	Init	72484	73175	+	0
	Intr	73451	73571	+	0
	Intr	73660	73844	+	0
60	Intr	73922	74053	+	0

2025 RELEASE UNDER E.O. 14176

1306

Term	74132	74268	+	0
------	-------	-------	---	---

>4063730 /38879

```
5      len =      2303      nex =      7
```

Term	91140	90858	-	0
------	-------	-------	---	---

Intr	91393	91280	-	0
------	-------	-------	---	---

Intr	91609	91538	-	0
------	-------	-------	---	---

10	Intr	92030	91961	-	0
----	------	-------	-------	---	---

Intr	92214	92117	-	0
------	-------	-------	---	---

Intr	92495	92294	-	0
------	-------	-------	---	---

Init	93160	92726	-	0
------	-------	-------	---	---

15 >4063735 /1351

```
len = 2263    nex = 3
```

```

Tnit 105458 105582      +      0

```

Init	105458	105982		
Int	106342	106839	+	0

Term	106936	107720	+	0
------	--------	--------	---	---

>4063735 /36464

```
25      len =      1399      nex =          2
```

```

Init  106315  106839      +      0

```

Term	106936	107713	+	0
------	--------	--------	---	---

30 >4063735 /110447

```
len =      659      nex =      1
```

Sngl	107104	107748	+	0
------	--------	--------	---	---

35 >4063735 /99323

```
len = 1093    nex = 1
```

40	Snq1	15861	16953	+	0
----	------	-------	-------	---	---

>4063735 /148671

```
len = 1079    nex = 2
```

Term	43823	43650	-	0
------	-------	-------	---	---

Term	44322	44347		0
Init	44402	44347	-	0

50 >4063735 /29766

```
len = 1949    nex = 6
```

Term	97324	96849	-	0
------	-------	-------	---	---

Inter	97547	97491	-	0
-------	-------	-------	---	---

55	Intr	97772	97630	-	0
----	------	-------	-------	---	---

Intr	98263	98128	-	0
------	-------	-------	---	---

Intr	98642	98446	-	0
------	-------	-------	---	---

Init	98797	98734	-	0
------	-------	-------	---	---

60 >4063737 /39796

1307

	len =	2050	nex =	8	
	Term	8947	8682	-	0
5	Intr	9161	9108	-	0
	Intr	9393	9318	-	0
	Intr	9521	9490	-	0
	Intr	9711	9616	-	0
	Intr	9843	9804	-	0
10	Intr	9992	9931	-	0
	Init	10722	10413	-	0
	>4063737	/37252			
15	len =	2858	nex =	12	
	Term	33875	33630	-	0
	Intr	34209	33988	-	0
	Intr	34433	34313	-	0
20	Intr	34722	34531	-	0
	Intr	34908	34836	-	0
	Intr	35127	35004	-	0
	Intr	35284	35208	-	0
	Intr	35494	35372	-	0
25	Intr	35672	35567	-	0
	Intr	35825	35763	-	0
	Intr	36084	35913	-	0
	Init	36487	36329	-	0
30	>4063737	/36243			
	len =	501	nex =	1	
	Sngl	56605	57105	+	0
35	>4063737	/13761			
	len =	2567	nex =	2	
40	Init	61983	62535	+	0
	Term	63541	64549	+	0
	>4063737	/17020			
45	len =	2217	nex =	2	
	Init	74589	74633	+	0
	Term	74734	75239	+	0
50	>4063756	/154687			
	len =	1056	nex =	3	
	Init	37451	37699	+	0
55	Intr	38169	38269	+	0
	Term	38354	38506	+	0
	>4063756	/15276			
60	len =	2125	nex =	4	

1308

	Init	37451	37699	+	0
	Intr	38169	38269	+	0
	Intr	38354	38547	+	0
5	Term	38976	39575	+	0
	>4079614 /9756				
	len =	974	nex =	1	
10	Sngl	27761	28734	+	0
	>4079614 /36980				
15	len =	1107	nex =	1	
	Sngl	38778	37672	-	0
	>4079614 /18761				
20	len =	250	nex =	1	
	Sngl	64013	63770	-	0
25	>4079614 /39763				
	len =	1030	nex =	1	
	Sngl	73340	74368	+	0
30	>4079614 /6208				
	len =	2548	nex =	5	
35	Init	75720	75893	+	0
	Intr	75986	76102	+	0
	Intr	76525	76695	+	0
	Intr	76777	76989	+	0
	Term	77081	77542	+	0
40	>4079614 /4512				
	len =	490	nex =	1	
45	Sngl	96496	96984	+	0
	>4092471 /22446				
	len =	506	nex =	1	
50	Sngl	11405	11769	+	0
	>4092471 /103691				
55	len =	1759	nex =	3	
	Term	42647	42351	-	0
	Intr	43888	43766	-	0
	Init	43998	43927	-	0
60					

ORIGINAL SUBMITTED

1309

>4092471 /20391

len = 1615 nex = 3

5	Term	42647	42495	-	0
	Intr	43888	43766	-	0
	Init	43998	43927	-	0

>4092472 /16473

10	len =	1615	nex =	1	
	Sngl	20093	20396	+	0

15 >4092472 /3126

len = 2865 nex = 5

20	Term	23958	23048	-	0
	Intr	24223	24110	-	0
	Intr	24606	24316	-	0
	Intr	24981	24673	-	0
	Init	25912	25060	-	0

25 >4092472 /104778

len = 3719 nex = 7

30	Term	46430	46119	-	0
	Intr	46578	46521	-	0
	Intr	46706	46670	-	0
	Intr	47519	47278	-	0
	Intr	47978	47862	-	0
	Intr	48234	48193	-	0
35	Init	49837	49522	-	0

>4096078 /155459

len = 415 nex = 1

40	Sngl	42191	42304	+	0
----	------	-------	-------	---	---

>4096078 /37034

45 len = 1948 nex = 10

	Term	45801	45724	-	0
	Intr	45988	45893	-	0
	Intr	46149	46063	-	0
50	Intr	46287	46219	-	0
	Intr	46556	46369	-	0
	Intr	46697	46640	-	0
	Intr	46828	46769	-	0
	Intr	46936	46901	-	0
55	Intr	47134	47023	-	0
	Init	47376	47303	-	0

>4096078 /21724

60 len = 2303 nex = 10

20250610 09:22:52

1310

	Term	45801	45724	-	0
	Intr	45988	45893	-	0
	Intr	46149	46063	-	0
5	Intr	46287	46219	-	0
	Intr	46556	46369	-	0
	Intr	46697	46640	-	0
	Intr	46828	46769	-	0
	Intr	46936	46901	-	0
10	Intr	47134	47023	-	0
	Init	47599	47303	-	0
	>4096078		/15956		
15	len =	1336	nex =	7	
	Init	54120	54197	+	0
	Intr	54285	54356	+	0
	Intr	54439	54543	+	0
20	Intr	54618	54710	+	0
	Intr	54798	54890	+	0
	Intr	54983	55106	+	0
	Term	55206	55327	+	0
25	>4096078		/8301		
	len =	1697	nex =	7	
	Term	55725	55556	-	0
30	Intr	55873	55814	-	0
	Intr	56063	55943	-	0
	Intr	56218	56142	-	0
	Intr	56400	56306	-	0
	Intr	56619	56513	-	0
35	Init	57252	56708	-	0
	>4096078		/20936		
	len =	1419	nex =	3	
40	Term	58407	58007	-	0
	Intr	58821	58569	-	0
	Init	59425	59245	-	0
45	>4096078		/11587		
	len =	698	nex =	1	
	Sngl	62604	61911	-	0
50	>4115352		/125955		
	len =	1433	nex =	3	
55	Term	26979	26916	-	0
	Intr	27312	27194	-	0
	Init	27672	27485	-	0
60	>4115352		/19338		

1311

	len =	1003	nex =	3	
	Term	26979	26681	-	0
	Intr	27312	27194	-	0
5	Init	27683	27485	-	0
>4115352 /41682					
10	len =	1450	nex =	2	
	Term	27312	27194	-	0
	Init	27690	27485	-	0
>4115370 /104717					
15	len =	1480	nex =	1	
	Sngl	21740	21519	-	0
20	>4115370 /117533				
	len =	1330	nex =	5	
	Term	43085	42601	-	0
25	Intr	43330	43160	-	0
	Intr	43471	43406	-	0
	Intr	43806	43554	-	0
	Init	43929	43880	-	0
30	>4115370 /40540				
	len =	2671	nex =	9	
	Term	86403	86022	-	0
35	Intr	86600	86484	-	0
	Intr	87059	86681	-	0
	Intr	87301	87146	-	0
	Intr	87515	87379	-	0
	Intr	87708	87589	-	0
40	Intr	87929	87801	-	0
	Intr	88216	88023	-	0
	Init	88692	88422	-	0
>4115912 /2343					
45	len =	2183	nex =	9	
	Term	15312	14953	-	0
	Intr	15468	15407	-	0
50	Intr	15691	15575	-	0
	Intr	15840	15781	-	0
	Intr	16048	15936	-	0
	Intr	16206	16134	-	0
	Intr	16322	16278	-	0
55	Intr	16522	16412	-	0
	Init	17135	16691	-	0
>4115912 /11730					
60	len =	896	nex =	2	

1312

	Init	34905	35153	+	0
	Term	35318	35800	+	0
5	>4115912 /8445				
	len =	1254	nex =	4	
	Init	95230	95390	+	0
10	Intr	95475	95562	+	0
	Intr	95755	95879	+	0
	Term	95956	96483	+	0
	>4115930 /40478				
15	len =	1543	nex =	2	
	Term	41954	41542	-	0
	Init	43084	42069	-	0
20	>4115930 /13513				
	len =	1185	nex =	4	
	Init	46052	46100	+	0
25	Intr	46383	46693	+	0
	Intr	46778	46829	+	0
	Term	47059	47236	+	0
30	>4115930 /39698				
	len =	1656	nex =	3	
	Init	53288	53526	+	0
35	Intr	54255	54411	+	0
	Term	54517	54943	+	0
	>4115930 /9218				
40	len =	1162	nex =	3	
	Init	59597	59737	+	0
	Intr	60190	60346	+	0
	Term	60446	60758	+	0
45	>4115930 /3033				
	len =	2350	nex =	8	
	Term	75735	75445	-	0
50	Intr	75948	75854	-	0
	Intr	76111	76037	-	0
	Intr	76336	76251	-	0
	Intr	76789	76738	-	0
55	Intr	77027	76953	-	0
	Intr	77177	77111	-	0
	Init	77787	77261	-	0
	>4159699 /1558				
60					

1313

	len =	1272	nex =	2	
	Term	14823	14303	-	0
	Init	15574	15206	-	0
5	>4159699 /17101				
	len =	2330	nex =	4	
10	Init	8693	8791	+	0
	Intr	9044	9107	+	0
	Intr	9186	9395	+	0
	Term	9532	10243	+	0
15	>4159700 /11277				
	len =	2170	nex =	11	
	Init	2785	3024	+	0
20	Intr	3119	3211	+	0
	Intr	3301	3401	+	0
	Intr	3481	3547	+	0
	Intr	3634	3733	+	0
	Intr	3811	3924	+	0
25	Intr	4060	4197	+	0
	Intr	4305	4389	+	0
	Intr	4488	4537	+	0
	Intr	4639	4703	+	0
	Term	4715	4949	+	0
30	>4159700 /9884				
	len =	1278	nex =	2	
35	Term	33490	33138	-	0
	Init	33894	33573	-	0
	>4159700 /106065				
40	len =	550	nex =	1	
	Sngl	36	578	+	0
	>4159700 /11225				
45	len =	1959	nex =	2	
	Init	5625	6382	+	0
	Term	6789	7074	+	0
50	>4159700 /37830				
	len =	2006	nex =	2	
55	Init	5116	6382	+	0
	Term	6789	7121	+	0
	>4159700 /39753				
60	len =	1990	nex =	2	

00100-942P-0000

1314

Init	5625	6382	+	0
Term	6789	7121	+	0

5 >4159702 /39782

len = 1769 nex = 4

Term	19118	18896	-	0
Intr	19546	19462	-	0
Intr	19766	19651	-	0
Init	20664	20383	-	0

>4159703 /22343

15 len = 2215 nex = 5

Init	7241	7578	+	0
Intr	8231	8293	+	0
Intr	8510	8587	+	0
Intr	8667	8829	+	0
Term	8902	9455	+	0

>4159704 /13704

25 len = 1335 nex = 4

Term	11416	11205	-	0
Intr	11620	11507	-	0
Intr	11891	11712	-	0
Init	12357	12143	-	0

>4159704 /21828

35 len = 1419 nex = 4

Term	11416	11151	-	0
Intr	11620	11507	-	0
Intr	11891	11712	-	0
Init	12357	12143	-	0

>4159704 /109598

45 len = 1100 nex = 2

Init	5940	6324	+	0
Term	6476	7039	+	0

>4159704 /39832

50 len = 2092 nex = 7

Init	71302	71572	+	0
Intr	71670	71791	+	0
Intr	71963	72067	+	0
Intr	72212	72304	+	0
Intr	72593	72655	+	0
Intr	72773	72882	+	0
Term	73080	73393	+	0

60

2025 RELEASE UNDER E.O. 14176

1315

>4159704 /13305

```
len = 2400    nex = 3
```

5	Init	80000	81012	+	0
	Intr	81347	81522	+	0
	Term	81623	82399	+	0

>4159705 /9508

```
len = 1639 nex = 3
```

	Init	28709	28753	+	0
	Intr	29227	29324	+	0
15	Term	29566	29821	+	0

>4159705 /114725

```
len = 490    nex = 1
```

Snq1	73649	73181	-	0
------	-------	-------	---	---

>4159705 /109298

```
25      len =      730      nex =      1
```

Snq1	73980	73255	-	0
------	-------	-------	---	---

>4159705 /326

```
len = 1716    nex = 1
```

Snq1	74641	73196	-	0
------	-------	-------	---	---

35 >4159705 /37994

```
len = 2303    nex = 7
```

40	Init	77530	77634	+	0
	Intr	77709	77959	+	0
	Intr	78103	78229	+	0
	Intr	78328	78601	+	0
	Intr	78834	78966	+	0
	Intr	79353	79439	+	0
45	Term	79694	79832	+	0

>4159706 /112601

```
len =      690      nex =      1
```

Snq1	10748	11437	+	0
------	-------	-------	---	---

>4159706 /16938

```
55      len =      1330      nex =          1
```

Snq1	14792	13472	-	0
------	-------	-------	---	---

>4159706 /30320

60

1316

	len =	370	nex =	1	
	Sngl	28894	28532	-	0
5	>4159706	/112749			
	len =	712	nex =	2	
10	Term	35876	35547	-	0
	Init	36258	35970	-	0
	>4159706	/8290			
15	len =	1320	nex =	4	
	Term	35876	35548	-	0
	Intr	36256	35970	-	0
	Intr	36422	36349	-	0
	Init	36867	36752	-	0
20	>4159706	/35999			
	len =	1933	nex =	7	
25	Init	53834	53966	+	0
	Intr	54115	54280	+	0
	Intr	54489	54541	+	0
	Intr	54877	54939	+	0
	Intr	55266	55316	+	0
30	Intr	55381	55498	+	0
	Term	55599	55766	+	0
	>4159706	/4581			
35	len =	1390	nex =	1	
	Sngl	58788	60171	+	0
	>4159706	/38468			
40	len =	1437	nex =	1	
	Sngl	65398	66834	+	0
45	>4159706	/35981			
	len =	1281	nex =	2	
50	Init	71773	71993	+	0
	Term	72232	73053	+	0
	>4159706	/27978			
55	len =	1451	nex =	5	
	Init	74975	75143	+	0
	Intr	75510	75632	+	0
	Intr	75728	76041	+	0
	Intr	76128	76180	+	0
60	Term	76272	76425	+	0

2025 RELEASE UNDER E.O. 14176

1317

5 >4159707 /207156
 len = 926 nex = 0
 10 >4159707 /41828
 len = 875 nex = 1
 Sngl 15648 15958 + 0
 15 >4159707 /29375
 len = 792 nex = 1
 Sngl 15648 15949 + 0
 20 >4159707 /37081
 len = 827 nex = 2
 Init 15176 15533 + 0
 Term 15648 16002 + 0
 25 >4159707 /18951
 len = 595 nex = 1
 Sngl 1560 2154 + 0
 30 >4159707 /32148
 len = 1343 nex = 1
 35 >4159707 /25313
 len = 1016 nex = 1
 Sngl 16466 17481 + 0
 >4159707 /41182
 45 len = 593 nex = 1
 Sngl 16874 17456 + 0
 50 >4159707 /28782
 len = 1096 nex = 3
 Init 27447 27699 + 0
 Intr 27790 28154 + 0
 Term 28252 28542 + 0
 60 >4159707 /97742
 len = 1220 nex = 2

1318

Init	2818	3153	+	0
Term	3582	4037	+	0

>4159707 /42151

5 len = 1186 nex = 2

Init	2848	3153	+	0
Term	3582	4033	+	0

10 >4159707 /2356

len = 2110 nex = 7

15	Term	28857	28597	-	0
	Intr	29050	28952	-	0
	Intr	29395	29294	-	0
	Intr	29617	29493	-	0
	Intr	29826	29753	-	0
20	Intr	30352	30211	-	0
	Init	30705	30436	-	0

>4159707 /21101

25 len = 310 nex = 1

Sngl 41123 40815 - 0

>4159707 /106867

30 len = 1340 nex = 2

	Term	41164	40724	-	0
	Init	41458	41253	-	0

35 >4159707 /22328

len = 1515 nex = 3

40	Term	41164	40703	-	0
	Intr	41458	41253	-	0
	Init	42217	42021	-	0

>4159707 /18636

45 len = 190 nex = 1

Sngl 42223 42042 - 0

50 >4159707 /38214

len = 1822 nex = 3

	Init	58012	58329	+	0
55	Intr	58492	58799	+	0
	Term	58885	59833	+	0

>4159707 /18979

60 len = 1890 nex = 3

1319

		Init	58012	58329	+	0
		Intr	58492	58799	+	0
		Term	58885	59901	+	0
5	>4159707		/26705			
	len =	538	nex =	1		
10	Sngl	63569	64106	+		0
	>4159707		/15303			
	len =	1493	nex =	2		
15	Term	68076	67448	-		0
	Init	68653	68151	-		0
	>4159707		/94231			
20	len =	1630	nex =	3		
	Init	78497	78952	+		0
	Intr	79546	79639	+		0
25	Term	79732	80120	+		0
	>4159708		/35429			
	len =	764	nex =	2		
30	Init	44910	45015	+		0
	Term	45368	45673	+		0
	>4159708		/16564			
35	len =	1359	nex =	5		
	Init	61816	62242	+		0
	Intr	62321	62370	+		0
40	Intr	62462	62549	+		0
	Intr	62641	62757	+		0
	Term	62943	63174	+		0
	>4159708		/39933			
45	len =	1197	nex =	1		
	Sngl	76529	77725	+		0
50	>4159709		/8209			
	len =	1632	nex =	3		
	Init	28608	28689	+		0
55	Intr	28882	28962	+		0
	Term	29255	29368	+		0
	>4159709		/20801			
60	len =	2026	nex =	8		

1320

	Init	28608	28689	+	0
	Intr	28882	28962	+	0
	Intr	29255	29368	+	0
5	Intr	29463	29564	+	0
	Intr	29876	29934	+	0
	Intr	30031	30133	+	0
	Intr	30228	30293	+	0
	Term	30387	30479	+	0
10	>4159709 /21868				
	len =	1990	nex =	3	
15	Term	1570	1249	-	0
	Intr	2549	2064	-	0
	Init	3231	2810	-	0
20	>4159709 /13261				
	len =	2987	nex =	3	
	Term	37003	36228	-	0
	Intr	38163	37674	-	0
25	Init	39214	38597	-	0
30	>4159709 /6579				
	len =	2352	nex =	8	
	Init	70836	71356	+	0
	Intr	71493	71667	+	0
	Intr	71764	71913	+	0
	Intr	72061	72222	+	0
35	Intr	72304	72489	+	0
	Intr	72587	72709	+	0
	Intr	72784	72873	+	0
	Term	72972	73187	+	0
40	>4159709 /9155				
	len =	557	nex =	1	
45	Sngl	75021	75577	+	0
	>4159709 /36834				
	len =	657	nex =	2	
50	Term	76162	75717	-	0
	Init	76373	76339	-	0
	>4159709 /115021				
55	len =	515	nex =	2	
	Init	82749	82901	+	0
	Term	83001	83063	+	0
60	>4159710 /13184				

2025-01-09 10:29:00

1321

	len =	1870	nex =	7	
	Term	14387	14080	-	0
5	Intr	14579	14479	-	0
	Intr	14816	14659	-	0
	Intr	14992	14930	-	0
	Intr	15289	15150	-	0
	Intr	15462	15369	-	0
10	Init	15947	15591	-	0
	>4159710 /100272				
	len =	670	nex =	1	
15	Sngl	18553	19215	+	0
	>4159711 /18503				
20	len =	1595	nex =	3	
	Term	15462	15206	-	0
	Intr	15812	15748	-	0
	Init	16800	16583	-	0
25	>4159712 /4044				
	len =	730	nex =	1	
30	Sngl	10754	10446	-	0
	>4159712 /36591				
	len =	2809	nex =	7	
35	Term	22587	22215	-	0
	Intr	22740	22669	-	0
	Intr	23033	22830	-	0
	Intr	23401	23272	-	0
40	Intr	23755	23529	-	0
	Intr	23937	23881	-	0
	Init	25023	24745	-	0
	>4159712 /14490				
45	len =	1090	nex =	3	
	Term	74186	73691	-	0
	Intr	74429	74331	-	0
50	Init	74774	74516	-	0
	>4159712 /1226				
	len =	794	nex =	2	
55	Init	8922	9122	+	0
	Term	9221	9715	+	0
	>4165340 /12055				
60					

1323

```
len =      397      nex =      1
```

Snq1	97179	97575	+	0
------	-------	-------	---	---

5 >4185120 /6390

```
len = 3432    nex = 13
```

Term	15226	14939	-	0
------	-------	-------	---	---

Intr	15428	15316	-	0
------	-------	-------	---	---

Intr	15588	15539	-	0
------	-------	-------	---	---

Intr	15779	15709	-	0
------	-------	-------	---	---

Intr	15948	15879	-	0
------	-------	-------	---	---

Intr	16280	16183	-	0
------	-------	-------	---	---

Intr	16473	16407	-	0
------	-------	-------	---	---

Tnt+r	16656	16579	-	0
-------	-------	-------	---	---

Intr	16926	16759	-	0
------	-------	-------	---	---

Intr	16926	16755		
Intr	17196	17029	-	0

Incl	17198	17829		
Trnr	17587	17291	-	0

	Incl	17587	17291		
20	Intr	17910	17665	-	0

Init	17910	17889		
Init	18370	18221	-	0

>4185120 /96856

```
25      len =      646      nex =      2
```

Term	42352	41925	-	0
------	-------	-------	---	---

Init	42557	42437	-	0
------	-------	-------	---	---

30 >4185120 /25812

```
len =      761    nex =      1
```

Sngl	99742	98982	-	0
------	-------	-------	---	---

35 >4185128 /92525

```
len = 1425    nex = 2
```

40	Init	13244	13883	+	0
----	------	-------	-------	---	---

Term	14344	14668	+	0
------	-------	-------	---	---

>4185128 /16416

```
45      len =      1248      nex =          1
```

Snq1	17454	18701	+	0
------	-------	-------	---	---

>4185128 /17464

50

Term	32604	32490	-	0
------	-------	-------	---	---

Tnt r	32905	32692	-	0
-------	-------	-------	---	---

55	Intr	33672	32987	-	0
----	------	-------	-------	---	---

Intr	33912	33755	-	0
------	-------	-------	---	---

Tnit	34329	34123	-	0
------	-------	-------	---	---

>4185128 /26663

60

1324

	len =	1076	nex =	4	
	Init	38740	39208	+	0
	Intr	39413	39478	+	0
5	Intr	39529	39651	+	0
	Term	39743	39815	+	0
	>4185128	/24619			
	len =	1711	nex =	5	
	Init	43094	43248	+	0
	Intr	43475	43537	+	0
	Intr	44259	44337	+	0
15	Intr	44425	44470	+	0
	Term	44574	44804	+	0
	>4185128	/14084			
	len =	1815	nex =	6	
	Init	54517	54995	+	0
	Intr	55165	55247	+	0
	Intr	55403	55492	+	0
25	Intr	55577	55725	+	0
	Intr	55817	55948	+	0
	Term	56044	56331	+	0
	>4185128	/4231			
	len =	1128	nex =	1	
	Sngl	57987	56860	-	0
35	>4185128	/37734			
	len =	1154	nex =	1	
	Sngl	7894	6748	-	0
40	>4191760	/35366			
	len =	2418	nex =	7	
45	Term	51175	50955	-	0
	Intr	51401	51262	-	0
	Intr	51669	51495	-	0
	Intr	51922	51764	-	0
	Intr	52401	52256	-	0
50	Intr	52598	52509	-	0
	Init	53372	52795	-	0
	>4191760	/16162			
55	len =	3435	nex =	7	
	Term	54020	53640	-	0
	Intr	54593	54540	-	0
	Intr	54849	54707	-	0
60	Intr	55352	54943	-	0

1325

Intr	55948	55819	-	0
Intr	56284	56184	-	0
Init	57074	56736	-	0

5 >4191760 /27782

len = 2185 nex = 6

Term	68405	68073	-	0
Intr	68552	68485	-	0
Intr	68822	68687	-	0
Intr	69025	68922	-	0
Intr	69410	69139	-	0
Init	70257	69851	-	0

15 >4191760 /31714

len = 550 nex = 2

Init	70697	70805	+	0
Term	71007	71062	+	0

>4191771 /2312

25 len = 2918 nex = 14

Term	28814	28356	-	0
Intr	28989	28901	-	0
Intr	29148	29085	-	0
Intr	29326	29244	-	0
Intr	29548	29422	-	0
Intr	29742	29658	-	0
Intr	29885	29822	-	0
Intr	30040	29979	-	0
Intr	30224	30154	-	0
Intr	30405	30310	-	0
Intr	30627	30479	-	0
Intr	30777	30708	-	0
Intr	30887	30847	-	0
Init	31273	30959	-	0

>4191771 /14853

45 len = 2857 nex = 12

Init	33231	33385	+	0
Intr	33490	33533	+	0
Intr	33762	33857	+	0
Intr	33987	34079	+	0
Intr	34302	34389	+	0
Intr	34494	34555	+	0
Intr	34638	34822	+	0
Intr	34906	35083	+	0
Intr	35171	35214	+	0
Intr	35298	35514	+	0
Intr	35611	35628	+	0
Term	35711	36087	+	0

60 >4191771 /35176

1326

	len =	2743	nex =	11	
	Init	33285	33385	+	0
	Intr	33490	33533	+	0
5	Intr	33762	33857	+	0
	Intr	33987	34079	+	0
	Intr	34302	34389	+	0
	Intr	34494	34555	+	0
10	Intr	34638	34822	+	0
	Intr	34906	35083	+	0
	Intr	35171	35214	+	0
	Intr	35298	35514	+	0
	Term	35711	36017	+	0
15	>4191771	/33083			
	len =	217	nex =	1	
20	Sngl	54044	54260	+	0
	>4191771	/125285			
	len =	447	nex =	1	
25	Sngl	65686	65240	-	0
	>4191771	/152285			
	len =	1045	nex =	1	
30	Sngl	66311	65267	-	0
	>4191771	/19555			
35	len =	1320	nex =	3	
	Term	71563	70726	-	0
	Intr	71884	71664	-	0
	Init	72045	71968	-	0
40	>4191771	/26452			
	len =	850	nex =	1	
45	Sngl	83459	83331	-	0
	>4199934	/29227			
	len =	1035	nex =	2	
50	Init	19693	19801	+	0
	Term	20093	20727	+	0
	>4199934	/33136			
55	len =	338	nex =	1	
	Sngl	20399	20729	+	0
60	>4199934	/154322			

1327

	len =	352	nex =	1	
5	Sngl	23117	23462	+	0
	>4199934 /36784				
	len =	1133	nex =	1	
10	Sngl	23117	24249	+	0
	>4204173 /41722				
	len =	1054	nex =	2	
15	Term	21811	21377	-	0
	Init	22430	21898	-	0
	>4204173 /14471				
20	len =	1120	nex =	3	
	Term	23568	23254	-	0
	Intr	23859	23729	-	0
25	Init	24373	23953	-	0
	>4217996 /8441				
	len =	1482	nex =	1	
30	Sngl	51080	52561	+	0
	>4217996 /38968				
35	len =	2388	nex =	9	
	Init	52825	53030	+	0
	Intr	53128	53176	+	0
	Intr	53267	53355	+	0
40	Intr	53445	53501	+	0
	Intr	53636	53698	+	0
	Intr	53788	53853	+	0
	Intr	53992	54078	+	0
	Intr	54220	54313	+	0
45	Term	54408	55212	+	0
	>4218109 /324				
	len =	1860	nex =	5	
50	Term	39406	38929	-	0
	Intr	39731	39476	-	0
	Intr	40077	39838	-	0
	Intr	40403	40276	-	0
55	Init	40788	40484	-	0
	>4218109 /1968				
60	len =	1690	nex =	5	

20250909 09:01:50

1328

	Init	42662	43143	+	0
	Intr	43303	43336	+	0
	Intr	43644	43748	+	0
	Intr	43827	44015	+	0
5	Term	44115	44344	+	0
	>4218109	/36080			
	len =	1657	nex =	4	
10	Term	53929	53688	-	0
	Intr	54126	54076	-	0
	Intr	54789	54211	-	0
	Init	55344	55008	-	0
15	>4218109	/38545			
	len =	654	nex =	2	
20	Term	89772	89411	-	0
	Init	90064	89883	-	0
	>4220468	/41875			
25	len =	1728	nex =	5	
	Term	25333	24750	-	0
	Intr	25490	25414	-	0
	Intr	25978	25767	-	0
30	Intr	26212	26057	-	0
	Init	26477	26297	-	0
	>4220468	/917			
35	len =	599	nex =	1	
	Sngl	26654	27252	+	0
	>4220468	/25454			
40	len =	310	nex =	1	
	Sngl	42788	42485	-	0
45	>4220468	/19157			
	len =	1214	nex =	1	
	Sngl	43720	42507	-	0
50	>4220468	/39244			
	len =	2268	nex =	0	
55	>4220468	/8247			
	len =	1486	nex =	5	
	Init	61354	61713	+	0
60	Intr	61815	61920	+	0

1329

Intr	62016	62099	+	0
Intr	62194	62322	+	0
Term	62414	62839	+	0

5 >4220468 /9469

len =	1120	nex =	5	
-------	------	-------	---	--

10

Term	75604	75323	-	0
Intr	75828	75680	-	0
Intr	76086	76037	-	0
Intr	76234	76176	-	0
Init	76442	76342	-	0

15 >4220468 /104853

len =	1161	nex =	3	
-------	------	-------	---	--

20

Term	80528	80150	-	0
Intr	80846	80610	-	0
Init	81310	80910	-	0

>4220510 /33382

25

len =	1810	nex =	2	
-------	------	-------	---	--

Term	119483	118560	-	0
Init	120362	119827	-	0

30 >4220510 /39461

len =	1133	nex =	2	
-------	------	-------	---	--

35

Term	62775	62278	-	0
Init	63410	63166	-	0

>4220510 /36681

40

len =	2001	nex =	3	
-------	------	-------	---	--

Init	8869	9090	+	0
Intr	9370	10119	+	0
Term	10211	10869	+	0

45 >4220627 /11357

len =	670	nex =	2	
-------	-----	-------	---	--

50

Term	23338	22980	-	0
Init	23642	23419	-	0

>4220627 /1267

55

len =	1798	nex =	2	
-------	------	-------	---	--

Init	30233	30321	+	0
Term	30435	32030	+	0

>4220630 /10008

60

2025 RELEASE UNDER E.O. 14176

1330

```
len =      2590      nex =          9

Term 16866 16556    -    0
Intr 17015 16956    -    0
Intr 17288 17193    -    0
Intr 18114 18032    -    0
Intr 18317 18255    -    0
Intr 18553 18457    -    0
Intr 18722 18641    -    0
Intr 18974 18880    -    0
Init 19138 19050    -    0
```

>4220630 / 33813

```
15      len =    1114    nex =         2
      Term    3819    3692    -         0
      Init    4805    4216    -         0
```

20 >4220631 /36799

len =	2843	nex =	10	
Init	27324	27510	+	0
Intr	27591	27698	+	0
Intr	27785	27877	+	0
Intr	27952	28044	+	0
Intr	28143	28228	+	0
Intr	28314	28592	+	0
Intr	28670	28803	+	0
Intr	28922	29014	+	0
Intr	29085	29202	+	0
Term	29308	29650	+	0

35 >4220631 /698

```
len =      550    nex =      1
Snq1 29083 29202      +      0
```

>4220631 /21717

len =	3328	nex =	8	
Init	32801	32924	+	0
Intr	33757	33803	+	0
Intr	34429	34757	+	0
Intr	34840	34922	+	0
Intr	35182	35257	+	0
Intr	35373	35437	+	0
Intr	35509	35604	+	0
Term	35716	36125	+	0

>4220631 /106086

```

55      len =      430      nex =      2
      Init 43315 43344      +      0
      Term 43425 43740      +      0

```

60

1331

>4220631 /4634

len = 271 nex = 1

5 Sngl 43471 43741 + 0

>4220631 /21618

len = 1544 nex = 1

10 Sngl 52445 52155 - 0

>4220632 /21608

15 len = 2002 nex = 6

Init 15759 16019 + 0

Intr 16186 16243 + 0

Intr 16326 16397 + 0

20 Intr 16993 17112 + 0

Intr 17201 17302 + 0

Term 17512 17760 + 0

>4220632 /33442

25 len = 323 nex = 1

Sngl 38225 38547 + 0

30 >4220632 /11562

len = 1630 nex = 1

Sngl 40558 40363 - 0

35 >4220632 /31276

len = 539 nex = 2

40 Term 42146 41798 - 0

Init 42336 42218 - 0

>4220633 /6602

45 len = 1537 nex = 2

Init 24986 25221 + 0

Term 25635 26107 + 0

50 >4220633 /14561

len = 1210 nex = 1

Sngl 41689 40482 - 0

55 >4220633 /43004

len = 825 nex = 1

60 Sngl 43944 44768 + 0

000102 02200000

1332

>4220633 /36013

5	len =	2510	nex =	11	
	Init	47139	47255	+	0
	Intr	47335	47437	+	0
	Intr	47521	47586	+	0
10	Intr	47676	47799	+	0
	Intr	47876	48010	+	0
	Intr	48100	48160	+	0
	Intr	48246	48322	+	0
	Intr	48458	48613	+	0
	Intr	48705	48833	+	0
15	Intr	48914	49046	+	0
	Term	49196	49300	+	0

>4220635 /11339

20	len =	2031	nex =	7	
	Init	38166	38351	+	0
	Intr	38520	38541	+	0
25	Intr	38891	39017	+	0
	Intr	39176	39266	+	0
	Intr	39361	39450	+	0
	Intr	39711	39811	+	0
	Term	39886	40196	+	0

>4220635 /144690

	len =	2028	nex =	7	
35	Init	38172	38351	+	0
	Intr	38520	38541	+	0
	Intr	38891	39017	+	0
	Intr	39176	39266	+	0
	Intr	39361	39450	+	0
40	Intr	39711	39811	+	0
	Term	39886	40199	+	0

>4220635 /7103

```

45      len =      626      nex =      1
      Snq1 40942 41567      +      0

```

>4220635 /25758

50	len =	1948	nex =	6	
	Init	44440	44586	+	0
	Intr	44775	44978	+	0
	Intr	45285	45530	+	0
55	Intr	45644	45886	+	0
	Intr	45969	46034	+	0
	Term	46128	46387	+	0

>4220635 /40185

60

1333

	len =	1376	nex =	1	
	Sngl	53066	54441	+	0
5	>4220636	/26820			
	len =	2132	nex =	10	
10	Init	22952	23071	+	0
	Intr	23266	23317	+	0
	Intr	23412	23607	+	0
	Intr	23692	23771	+	0
	Intr	23869	23985	+	0
15	Intr	24090	24140	+	0
	Intr	24233	24299	+	0
	Intr	24382	24485	+	0
	Intr	24578	24682	+	0
	Term	24783	25083	+	0
20	>4220636	/121993			
	len =	1306	nex =	5	
25	Term	37609	37394	-	0
	Intr	38097	37965	-	0
	Intr	38318	38196	-	0
	Intr	38425	38399	-	0
	Init	38699	38603	-	0
30	>4220637	/29678			
	len =	1417	nex =	2	
35	Term	13304	12526	-	0
	Init	13942	13414	-	0
	>4220637	/19279			
40	len =	614	nex =	1	
	Sngl	17264	17877	+	0
	>4220637	/26411			
45	len =	1523	nex =	4	
	Init	36910	37153	+	0
	Intr	37747	37816	+	0
50	Intr	37913	38021	+	0
	Term	38124	38432	+	0
	>4220637	/5961			
55	len =	1390	nex =	4	
	Init	38657	38747	+	0
	Intr	39193	39308	+	0
	Intr	39403	39466	+	0
60	Term	39552	40045	+	0

1334

>4220637 /32574

len = 2029 nex = 4

5	Init	52862	53196	+	0
	Intr	53299	53516	+	0
	Intr	53613	53860	+	0
	Term	53997	54890	+	0

10 >4220638 /36765

len = 1180 nex = 3

15	Term	10773	10592	-	0
	Intr	11443	10925	-	0
	Init	11771	11524	-	0

>4220638 /16377

20 len = 1782 nex = 4

	Term	10773	10360	-	0
	Intr	11443	10925	-	0
	Intr	11826	11524	-	0
25	Init	12141	11914	-	0

>4220638 /7015

30 len = 586 nex = 2

	Init	30811	30933	+	0
	Term	31055	31396	+	0

>4220638 /19110

35 len = 1340 nex = 2

	Term	33848	32699	-	0
	Init	34038	33944	-	0

40 >4220638 /102368

len = 730 nex = 2

45	Init	40029	40250	+	0
	Term	40282	40754	+	0

>4220640 /32066

50 len = 875 nex = 2

	Term	778	350	-	0
	Init	1099	1015	-	0

55 >4220640 /101893

len = 646 nex = 1

60	Sngl	1489	2131	+	0
----	------	------	------	---	---

1335

>4220640 /6915

len = 2050 nex = 6

5	Init	21763	22063	+	0
	Intr	22330	22386	+	0
	Intr	22578	22783	+	0
	Intr	22894	23167	+	0
	Intr	23311	23382	+	0
10	Term	23473	23804	+	0

>4220640 /28838

len = 594 nex = 1

15	Sngl	27690	27097	-	0
----	------	-------	-------	---	---

>4220640 /33726

20	len =	1839	nex =	0
----	-------	------	-------	---

>4220640 /37012

25	len =	441	nex =	1	
	Sngl	31588	32028	+	0

>4220640 /20116

30	len =	1516	nex =	4	
	Term	32430	31988	-	0
	Intr	32676	32525	-	0
	Intr	33332	32969	-	0
35	Init	33503	33404	-	0

>4220640 /6156

40	len =	1498	nex =	3	
	Term	32430	32015	-	0
	Intr	32676	32525	-	0
	Init	33512	32969	-	0

45 >4220640 /10926

len = 1533 nex = 4

50	Term	32430	32015	-	0
	Intr	32676	32525	-	0
	Intr	33332	32969	-	0
	Init	33547	33404	-	0

>4220640 /2767

55	len =	1053	nex =	3	
	Term	34226	34114	-	0
	Intr	34482	34391	-	0
60	Init	34843	34602	-	0

2025 RELEASE UNDER E.O. 14176

>4220640 /102455

5	len =	1393	nex =	5	
	Init	36025	36150	+	0
	Intr	36243	36283	+	0
	Intr	36497	36565	+	0
	Intr	36651	36707	+	0
10	Term	36797	37018	+	0

>4220640 /6800

15	len =	1697	nex =	5	
	Term	45977	45853	-	0
	Intr	46279	46055	-	0
	Intr	46510	46382	-	0
	Intr	46828	46790	-	0
20	Init	47542	47110	-	0

>4220640 /141813

25	len =	951	nex =	1	
	Sngl	5195	4245	-	0

>4220640 /40646

30	len =	3764	nex =	4	
	Init	65692	65939	+	0
	Intr	66421	66558	+	0
	Intr	66651	66740	+	0
35	Term	66834	66890	+	0

>4220640 /28790

40	len =	1814	nex =	8	
	Init	71705	71840	+	0
	Intr	71931	71971	+	0
	Intr	72059	72313	+	0
	Intr	72390	72450	+	0
45	Intr	72546	72670	+	0
	Intr	72752	72814	+	0
	Intr	72899	73023	+	0
	Term	73185	73518	+	0

50 >4220640 /10103

	len =	610	nex =	1	
	Sngl	82414	83015	+	0

55 >4220641 /41308

	len =	1514	nex =	6	
60	Term	16718	16586	-	0

2025 RELEASE UNDER E.O. 14176

1337

	Intr	16848	16803	-	0
	Intr	17051	16951	-	0
	Intr	17286	17149	-	0
	Intr	17585	17487	-	0
5	Init	17849	17733	-	0
>4220641		/11265			
10	len =	1899	nex =	6	
	Term	28923	28663	-	0
	Intr	29185	29124	-	0
	Intr	29344	29270	-	0
	Intr	29971	29425	-	0
15	Intr	30137	30073	-	0
	Init	30561	30314	-	0
>4220641		/1152			
20	len =	1189	nex =	3	
	Term	68284	68140	-	0
	Intr	68713	68513	-	0
	Init	69328	69078	-	0
25	>4220643 /125902				
	len =	550	nex =	1	
30	Sngl	11332	11876	+	0
>4220643		/8788			
35	len =	1231	nex =	3	
	Term	12170	11960	-	0
	Intr	12444	12249	-	0
	Init	13190	12769	-	0
40	>4220643 /10876				
	len =	1876	nex =	6	
	Init	13443	13678	+	0
45	Intr	13778	13880	+	0
	Intr	14150	14217	+	0
	Intr	14462	14653	+	0
	Intr	14738	14856	+	0
	Term	14959	15318	+	0
50	>4220643 /109639				
	len =	2710	nex =	5	
55	Term	18433	18225	-	0
	Intr	20036	19936	-	0
	Intr	20200	20123	-	0
	Intr	20322	20279	-	0
	Init	20572	20403	-	0
60					

2025-09-25 09:40:40

1338

>4220643 /92267

	len =	1549	nex =	6	
5	Term	27866	27561	-	0
	Intr	28045	27980	-	0
	Intr	28239	28143	-	0
	Intr	28564	28318	-	0
	Intr	28840	28748	-	0
10	Init	29109	28937	-	0

>4220643 /2347

	len =	833	nex =	1	
15	Sngl	35179	34347	-	0

>4220643 /10875

20	len =	2193	nex =	6	
	Term	2953	2902	-	0
	Intr	3203	3096	-	0
	Intr	3434	3291	-	0
25	Intr	3609	3517	-	0
	Intr	3792	3721	-	0
	Init	4031	3873	-	0

>4220643 /126592

30	len =	2377	nex =	6	
	Term	7320	7059	-	0
	Intr	7755	7637	-	0
35	Intr	8012	7841	-	0
	Intr	8587	8399	-	0
	Intr	9007	8930	-	0
	Init	9247	9210	-	0

40 >4220644 /40534

	len =	2331	nex =	6	
	Term	14665	14123	-	0
45	Intr	15328	14749	-	0
	Intr	15498	15413	-	0
	Intr	15667	15597	-	0
	Intr	15888	15738	-	0
	Init	16453	16250	-	0

50 >4220644 /39041

	len =	1930	nex =	5	
55	Init	17880	18258	+	0
	Intr	18361	18441	+	0
	Intr	18898	19002	+	0
	Intr	19139	19288	+	0
	Term	19396	19805	+	0

60

000100-001000

1339

>4220644 /12996

len = 2758 nex = 8

5	Init	3859	4029	+	0
	Intr	4630	4933	+	0
	Intr	5050	5142	+	0
	Intr	5216	5349	+	0
	Intr	5438	5564	+	0
10	Intr	5941	6046	+	0
	Intr	6174	6245	+	0
	Term	6344	6616	+	0

>4220644 /40770

15 len = 2268 nex = 10

	Term	60836	60743	-	0
	Intr	60995	60915	-	0
20	Intr	61151	61082	-	0
	Intr	61384	61296	-	0
	Intr	61819	61711	-	0
	Intr	62039	61921	-	0
	Intr	62183	62124	-	0
25	Intr	62462	62339	-	0
	Intr	62674	62574	-	0
	Init	63010	62903	-	0

>4220644 /37152

30 len = 2246 nex = 5

	Init	68479	69417	+	0
	Intr	69503	69580	+	0
35	Intr	69662	69773	+	0
	Intr	70259	70309	+	0
	Term	70434	70724	+	0

>4220645 /19689

40 len = 1479 nex = 3

	Init	18327	18524	+	0
	Intr	18628	18878	+	0
45	Term	19278	19805	+	0

>4220645 /102797

len = 1492 nex = 0

50 >4220645 /40166

len = 430 nex = 1

55 Sngl 19392 19821 + 0

>4220645 /8735

len = 1330 nex = 6

60

1340

	Term	20047	19864	-	0
	Intr	20212	20150	-	0
	Intr	20365	20303	-	0
	Intr	20513	20457	-	0
5	Intr	20721	20611	-	0
	Init	21190	20957	-	0
>4220645		/29889			
10	len =	1879	nex =	3	
	Init	24415	24465	+	0
	Intr	24561	25553	+	0
	Term	25635	26068	+	0
15	>4220645		/24775		
	len =	1521	nex =	5	
20	Term	2820	2592	-	0
	Intr	3077	2961	-	0
	Intr	3417	3180	-	0
	Intr	3923	3810	-	0
	Init	4112	4012	-	0
25	>4220645		/8347		
	len =	531	nex =	1	
30	Sngl	53094	52564	-	0
	>4220645		/30517		
	len =	1717	nex =	5	
35	Init	69636	70034	+	0
	Intr	70115	70245	+	0
	Intr	70339	70572	+	0
	Intr	70658	70913	+	0
40	Term	71018	71352	+	0
>4220645		/13425			
	len =	573	nex =	2	
45	Init	70771	70913	+	0
	Term	71018	71340	+	0
>4235150		/15739			
50	len =	894	nex =	1	
	Sngl	33125	32244	-	0
55	>4235150		/21075		
	len =	811	nex =	2	
	Init	52126	52284	+	0
60	Term	52598	52936	+	0

ORIGINAL SUBMITTED

1341

>4235150 /110090

5	len =	778	nex =	2	
	Init	52159	52284	+	0
	Term	52598	52936	+	0

>4235150 /30523

10	len =	537	nex =	1	
	Sngl	54031	54567	+	0

15 >4235150 /32316

	len =	521	nex =	1	
20	Sngl	56193	56713	+	0

>4235150 /111638

	len =	1210	nex =	3	
25	Term	71107	70335	-	0
	Intr	71256	71195	-	0
	Init	71543	71346	-	0

>4235150 /32272

30	len =	1224	nex =	3	
	Term	71107	70354	-	0
	Intr	71256	71195	-	0
35	Init	71577	71346	-	0

>4235150 /25124

40	len =	1278	nex =	3	
	Term	71107	70354	-	0
	Intr	71256	71195	-	0
	Init	71631	71346	-	0

45 >4249393 /6819

	len =	1766	nex =	2	
50	Term	43545	42141	-	0
	Init	43906	43634	-	0

>4249393 /6017

55	len =	2858	nex =	14	
	Init	47446	47702	+	0
	Intr	47789	47851	+	0
	Intr	47928	48017	+	0
	Intr	48101	48160	+	0
60	Intr	48246	48311	+	0

2025 RELEASE UNDER E.O. 14176

1342

	Intr	48386	48451	+	0
	Intr	48684	48748	+	0
	Intr	48827	48919	+	0
	Intr	49182	49266	+	0
5	Intr	49347	49390	+	0
	Intr	49467	49529	+	0
	Intr	49655	49708	+	0
	Intr	49773	49832	+	0
	Term	49904	50303	+	0
10	>4249393 /8332				
	len =	670	nex =	3	
15	Init	49619	49708	+	0
	Intr	49773	49832	+	0
	Term	49904	50282	+	0
20	>4249393 /38690				
	len =	2474	nex =	4	
	Term	51247	50779	-	0
	Intr	51460	51331	-	0
25	Intr	51783	51576	-	0
	Init	53252	52424	-	0
	>4249393 /34081				
30	len =	1122	nex =	4	
	Init	65982	66175	+	0
	Intr	66322	66532	+	0
	Intr	66615	66784	+	0
35	Term	66874	67103	+	0
	>4249393 /39971				
40	len =	655	nex =	1	
	Sngl	72891	72237	-	0
	>4262221 /40457				
45	len =	532	nex =	1	
	Sngl	10235	10766	+	0
50	>4262221 /19163				
	len =	1709	nex =	2	
	Term	17710	16611	-	0
	Init	18319	18101	-	0
55	>4262221 /17996				
	len =	2650	nex =	2	
60	Init	24589	25919	+	0

2025 RELEASE UNDER E.O. 14176

1343

	Term	26031	27234	+	0
	>4262221 /94462				
5	len =	1422	nex =	6	
	Term	42754	42345	-	0
	Intr	42956	42854	-	0
	Intr	43130	43051	-	0
10	Intr	43426	43369	-	0
	Intr	43616	43519	-	0
	Init	43766	43687	-	0
	>4262221 /108290				
15	len =	1635	nex =	6	
	Term	42754	42345	-	0
	Intr	42956	42854	-	0
20	Intr	43130	43051	-	0
	Intr	43426	43369	-	0
	Intr	43616	43519	-	0
	Init	43869	43687	-	0
25	>4262221 /8949				
	len =	2493	nex =	6	
	Term	45313	44902	-	0
30	Intr	45683	45592	-	0
	Intr	45851	45766	-	0
	Intr	46447	46219	-	0
	Intr	46760	46534	-	0
	Init	47394	46849	-	0
35	>4262221 /3831				
	len =	1431	nex =	1	
40	Sngl	59938	58508	-	0
	>4262221 /120459				
	len =	670	nex =	1	
45	Sngl	61325	60659	-	0
	>4262221 /22418				
50	len =	1779	nex =	4	
	Term	5241	4407	-	0
	Intr	5454	5345	-	0
	Intr	5819	5550	-	0
55	Init	6185	5932	-	0
	>4262221 /37962				
60	len =	1312	nex =	3	

1344

Init	73550	73783	+	0
Intr	73889	74305	+	0
Term	74608	74861	+	0

5 >4262221 /2877

len = 2110 nex = 7

Term	75024	74744	-	0
Intr	75175	75108	-	0
Intr	75424	75248	-	0
Intr	75663	75511	-	0
Intr	75879	75796	-	0
Intr	76233	76087	-	0
Init	76452	76319	-	0

>4263038 /95374

20 len = 430 nex = 1

Sngl 43068 43493 + 0

>4263038 /3879

25 len = 642 nex = 1

Sngl 43783 43142 - 0

>4263373 /255

30 len = 7618 nex = 4

Term	38456	36387	-	0
Intr	40079	40029	-	0
Intr	43641	40959	-	0
Init	44004	43817	-	0

>4263373 /4059

40 len = 7616 nex = 4

Term	38456	36389	-	0
Intr	40079	40029	-	0
Intr	43641	40959	-	0
Init	44004	43817	-	0

>4263373 /42686

50 len = 1090 nex = 2

Term	58099	57610	-	0
Init	58697	58409	-	0

>4263373 /28019

55 len = 1179 nex = 2

Term	58099	57546	-	0
Init	58724	58409	-	0

60

2025 RELEASE UNDER E.O. 14176

1345

>4263586 /2771

	len =	850	nex =	5	
5	Init	44628	44729	+	0
	Intr	44822	44985	+	0
	Intr	45066	45099	+	0
	Intr	45171	45264	+	0
	Term	45352	45476	+	0

10 >4263586 /42384

	len =	868	nex =	5	
15	Init	44631	44729	+	0
	Intr	44822	44985	+	0
	Intr	45066	45099	+	0
	Intr	45171	45264	+	0
	Term	45352	45498	+	0

20 >4263586 /34420

	len =	1907	nex =	8	
25	Term	54050	53865	-	0
	Intr	54344	54138	-	0
	Intr	54536	54429	-	0
	Intr	54749	54633	-	0
	Intr	54915	54829	-	0
30	Intr	55154	54990	-	0
	Intr	55337	55226	-	0
	Init	55771	55407	-	0

35 >4263586 /1237

	len =	2153	nex =	8	
	Init	55911	56299	+	0
	Intr	56483	56550	+	0
40	Intr	56661	56794	+	0
	Intr	56893	56981	+	0
	Intr	57082	57139	+	0
	Intr	57205	57257	+	0
	Intr	57350	57562	+	0
45	Term	57841	58063	+	0

>4263694 /19875

	len =	171	nex =	1	
50	Sngl	10500	10333	-	0

>4263694 /31665

55	len =	733	nex =	3	
	Term	10901	10692	-	0
	Intr	11155	11017	-	0
	Init	11424	11303	-	0

60

2025 RELEASE UNDER E.O. 14176

1346

>4263694 /40344

len = 701 nex = 2

5	Term	14673	14535	-	0
	Init	14888	14792	-	0

>4263694 /159279

10	len =	746	nex =	1	
	Sngl	16225	16970	+	0

>4263694 /32907

15	len =	790	nex =	3	
	Term	29067	28667	-	0
	Intr	29226	29161	-	0
20	Init	29452	29310	-	0

>4263694 /37176

25	len =	1706	nex =	7	
	Term	29067	28646	-	0
	Intr	29226	29161	-	0
	Intr	29522	29310	-	0
	Intr	29704	29617	-	0
30	Intr	29841	29804	-	0
	Intr	30029	29932	-	0
	Init	30183	30118	-	0

>4263694 /35952

35	len =	1351	nex =	4	
	Term	43833	43628	-	0
	Intr	44147	44072	-	0
40	Intr	44489	44259	-	0
	Init	44978	44842	-	0

>4263694 /13834

45	len =	1330	nex =	5	
	Init	49040	49096	+	0
	Intr	49251	49313	+	0
	Intr	49548	49848	+	0
50	Intr	49937	50002	+	0
	Term	50123	50363	+	0

>4263694 /13453

55	len =	1270	nex =	4	
	Init	49251	49313	+	0
	Intr	49548	49848	+	0
	Intr	49937	50002	+	0
60	Term	50123	50363	+	0

1347

>4263694 /9338

5	len =	754	nex =	3	
	Init	49575	49848	+	0
	Intr	49937	50002	+	0
	Term	50123	50328	+	0

10 >4263694 /15582

	len =	2721	nex =	9	
	Term	52117	51580	-	0
15	Intr	52340	52248	-	0
	Intr	52602	52434	-	0
	Intr	52976	52716	-	0
	Intr	53226	53078	-	0
	Intr	53488	53326	-	0
20	Intr	53624	53560	-	0
	Intr	53954	53704	-	0
	Init	54300	54063	-	0

25 >4263694 /21201

	len =	707	nex =	1	
	Sngl	57642	58348	+	0

30 >4263694 /35447

	len =	804	nex =	2	
	Init	79508	79806	+	0
35	Term	79888	80311	+	0

>4263694 /37871

40	len =	460	nex =	2	
	Init	79509	79806	+	0
	Term	79888	79968	+	0

>4263753 /29201

45	len =	1898	nex =	5	
	Term	23901	23530	-	0
	Intr	24129	24008	-	0
50	Intr	24602	24242	-	0
	Intr	25041	24923	-	0
	Init	25424	25374	-	0

>4263753 /38824

55	len =	2017	nex =	5	
	Term	23901	23444	-	0
	Intr	24129	24008	-	0
60	Intr	24602	24242	-	0

1348

Intr	25041	24923	-	0
Init	25447	25374	-	0

>4263753 /32341

5	len =	1853	nex =	4
---	-------	------	-------	---

Term	25964	25590	-	0
Intr	26181	26057	-	0
Intr	26645	26279	-	0
Init	27442	27181	-	0

>4263753 /18342

15	len =	1999	nex =	3
----	-------	------	-------	---

Term	25964	25587	-	0
Intr	26181	26057	-	0
Init	26645	26279	-	0

>4263753 /142593

len =	2369	nex =	9
-------	------	-------	---

25	Term	27931	27684	-	0
	Intr	28203	28139	-	0
	Intr	28363	28298	-	0
	Intr	28872	28453	-	0
	Intr	29202	29080	-	0
30	Intr	29421	29303	-	0
	Intr	29591	29498	-	0
	Intr	29825	29715	-	0
	Init	30052	29909	-	0

>4263753 /114575

len =	2332	nex =	3
-------	------	-------	---

40	Term	40874	40258	-	0
	Intr	42140	40970	-	0
	Init	42589	42333	-	0

>4263753 /104891

45	len =	1090	nex =	1	
	Sngl	49712	49632	-	0

>4263753 /34854

50	len =	771	nex =	0
----	-------	-----	-------	---

>4263753 /7959

55	len =	1461	nex =	1	
	Sngl	49712	49320	-	0

>4263753 /22013

60

20250901 09:01:00

1349

	len =	1473	nex =	3	
	Init	59531	59715	+	0
	Intr	60399	60450	+	0
5	Term	60528	61003	+	0
	>4263753 /121365				
10	len =	1415	nex =	2	
	Init	59581	59715	+	0
	Term	60399	60450	+	0
	>4263753 /8599				
15	len =	1015	nex =	1	
	Sngl	72614	72191	-	0
20	>4263753 /42931				
	len =	2493	nex =	4	
	Term	72614	72054	-	0
25	Intr	73868	73162	-	0
	Intr	74336	74237	-	0
	Init	74546	74430	-	0
	>4263753 /33691				
30	len =	2530	nex =	4	
	Term	72614	72054	-	0
	Intr	73868	73162	-	0
35	Intr	74336	74237	-	0
	Init	74574	74430	-	0
	>4263762 /24286				
40	len =	2117	nex =	3	
	Init	71346	71674	+	0
	Intr	71760	71795	+	0
	Term	73186	73462	+	0
45	>4263774 /9149				
	len =	1339	nex =	3	
50	Term	11294	10832	-	0
	Intr	11773	11641	-	0
	Init	12170	11858	-	0
	>4263774 /8134				
55	len =	1118	nex =	2	
	Term	959	793	-	0
	Init	1290	1112	-	0
60					

1350

>4263774 /112110

len = 1307 nex = 2

5	Term	959	618	-	0
	Init	1290	1112	-	0

>4263774 /5598

10 len = 1055 nex = 2

	Init	22583	22724	+	0
	Term	23358	23637	+	0

15 >4263774 /9996

len = 571 nex = 1

20 Sngl 29990 30560 + 0

>4263774 /22447

len = 1408 nex = 2

25	Term	2804	2016	-	0
	Init	3423	3330	-	0

>4263774 /40628

30 len = 2445 nex = 9

	Init	4021	4390	+	0
	Intr	4472	4567	+	0
	Intr	4643	4735	+	0
35	Intr	4808	4881	+	0
	Intr	4958	5224	+	0
	Intr	5677	5804	+	0
	Intr	5890	5985	+	0
40	Intr	6073	6178	+	0
	Term	6264	6465	+	0

>4263774 /38267

45 len = 2290 nex = 6

	Init	45101	45783	+	0
	Intr	45981	46103	+	0
	Intr	46189	46282	+	0
	Intr	46402	46484	+	0
50	Intr	46569	46658	+	0
	Term	46747	47387	+	0

>4263774 /2403

55 len = 1055 nex = 3

	Term	62305	62089	-	0
	Intr	62745	62486	-	0
	Init	63143	63001	-	0

60

ORIGINAL PAGE 500

1351

>4309719 /38985

```
len =      748    nex =      2
```

5	Term	21312	20917	-	0
	Init	21664	21527	-	0

>4309719 /692

```
10      len =      1100      nex =          2
```

Init	24887	24997	+	0
Term	25486	25986	+	0

15 >4309719 /5695

```
len = 1349    nex = 5
```

20	Term	26228	25952	-	0
	Intr	26479	26399	-	0
	Intr	26755	26593	-	0
	Intr	26979	26869	-	0
	Init	27300	27066	-	0

25 >4309719 /105626

```
len = 1111 nex = 2
```

	Term	33999	33726	-	0
30	Init	34836	34438	-	0

>4309719 /4845

```
len = 1063 nex = 1
```

Sngl	43384	44446	+	0
------	-------	-------	---	---

>4309719 /2731

```
40      len =      971      nex =      3
```

Init	60881	61402	+	0
Intr	61481	61699	+	0
Term	61778	61851	+	0

45
 >4309719 /32047

```
len = 1818    nex = 6
```

50	Init	65313	65489	+	0
	Intr	65653	65708	+	0
	Intr	65790	65882	+	0
	Intr	65980	66050	+	0
	Intr	66683	66787	+	0
55	Term	66874	67130	+	0

>4309719 /8068

```
len = 2380 nex = 8
```

1352

		Init	81136	81254	+	0
		Intr	82137	82165	+	0
		Intr	82264	82329	+	0
5		Intr	82552	82659	+	0
		Intr	82754	82819	+	0
		Intr	82909	83029	+	0
		Intr	83121	83202	+	0
		Term	83284	83515	+	0
10	>4309719		/150251			
	len =	550	nex =	2		
15		Init	83121	83202	+	0
		Term	83284	83538	+	0
	>4309719		/35743			
20	len =	1049	nex =	1		
	Sngl	89585	90633	+	0	
	>4309747		/34060			
25	len =	1132	nex =	5		
	Init	38591	38699	+	0	
	Intr	38809	38862	+	0	
30	Intr	39074	39183	+	0	
	Intr	39346	39430	+	0	
	Term	39528	39722	+	0	
	>4314354		/1650			
35	len =	1096	nex =	2		
	Init	33624	33916	+	0	
	Term	34315	34719	+	0	
40	>4314354		/19093			
	len =	1161	nex =	5		
	Term	36756	36705	-	0	
45	Intr	36980	36836	-	0	
	Intr	37153	37045	-	0	
	Intr	37293	37236	-	0	
	Init	37466	37365	-	0	
50	>4314354		/107116			
	len =	379	nex =	1		
	Sngl	41136	41514	+	0	
55	>4314354		/37274			
	len =	2230	nex =	6		
60	Init	46941	47138	+		

1353

	Intr	47607	48124	+	0
	Intr	48222	48451	+	0
	Intr	48548	48689	+	0
	Intr	48776	48837	+	0
5	Term	48917	49168	+	0
	>4314354 /4014				
10	len =	1725	nex =	1	
	Sngl	61481	60193	-	0
	>4314374 /18569				
15	len =	1010	nex =	3	
	Term	443	32	-	0
	Intr	853	747	-	0
20	Init	1041	1020	-	0
	>4314374 /38329				
	len =	2442	nex =	8	
25	Init	2646	2694	+	0
	Intr	2782	2852	+	0
	Intr	3021	3085	+	0
	Intr	3181	3225	+	0
	Intr	3316	3396	+	0
30	Intr	3473	3540	+	0
	Intr	3616	3664	+	0
	Term	3771	3925	+	0
	>4314374 /35081				
35	len =	839	nex =	2	
	Term	22049	21488	-	0
40	Init	22326	22150	-	0
	>4314374 /20875				
	len =	2177	nex =	9	
45	Term	42829	42511	-	0
	Intr	43040	42944	-	0
	Intr	43426	43260	-	0
	Intr	43549	43516	-	0
	Intr	43684	43638	-	0
50	Intr	43912	43761	-	0
	Intr	44106	44015	-	0
	Intr	44262	44214	-	0
	Init	44687	44480	-	0
55	>4314374 /1126				
	len =	1671	nex =	3	
	Term	57444	57236	-	0
60	Intr	58681	57651	-	0

2025 RELEASE UNDER E.O. 14176

Reference No. 2750-942P

	Init	58906	58778	-	1354 0
	>4314374	/26333			
5	len =	1393	nex =	0	
	>4325340	/27144			
10	len =	460	nex =	1	
	Sngl	950	630	-	0
	>4325340	/19541			
15	len =	670	nex =	1	
	Sngl	950	636	-	0
	>4325340	/122423			
20	len =	624	nex =	1	
	Sngl	30231	30854	+	0
25	>4325352	/9742			
	len =	131	nex =	1	
	Sngl	51030	51160	+	0
30	>4325365	/33985			
	len =	2158	nex =	6	
35	Term	346	1	-	0
	Intr	977	867	-	0
	Intr	1268	1059	-	0
	Intr	1516	1383	-	0
	Intr	1864	1705	-	0
40	Init	2158	1967	-	0
	>4325365	/35963			
	len =	946	nex =	2	
45	Init	65779	66009	+	0
	Term	66230	66707	+	0
	>4325365	/3552			
50	len =	1150	nex =	1	
	Sngl	93731	94876	+	0
55	>4325365	/6659			
	len =	1125	nex =	1	
	Sngl	93738	94862	+	0
60					

1355

>4325365 /36687

```
len = 1990    nex = 1
```

5 Sngl 99721 100926 + 0

>4335711 /5

```
len = 1270    nex = 1
```

10					
	Sngl	115475	114208	-	0

>4335711 /9064

```
15      len =      650      nex =      0
```

>4335711 /7216

```
len = 1048    nex = 1
```

20				
	Sng1	35552	34505	- 0

>4335711 /7792

```
25      len =      1193      nex =          3
```

Init	46238	46470	+	0
------	-------	-------	---	---

Intr	46624	46739	+	0
------	-------	-------	---	---

Term	47125	47430	+	0
------	-------	-------	---	---

30
>4335711 /122361

```
len =      718      nex =      2
```

```
35      Init    46239    46470      +      0
```

Term	46624	46956	+	0
------	-------	-------	---	---

>4335711 /7201

```
40      len =      1192      nex =      3
```

Init	46239	46470	+	0
------	-------	-------	---	---

Intr	46624	46739	+	0
------	-------	-------	---	---

Term	47125	47430	+	0
------	-------	-------	---	---

45
>4335711 /40916

```
len =      1194      nex =          3
```

```
50      Init    46239    46470      +      0
```

Intr	46624	46739	+	0
------	-------	-------	---	---

Term	47125	47432	+	0
------	-------	-------	---	---

>4335711 /9723

```
55      len =      2068      nex =      1
```

Snq1	82273	80206	-	0
------	-------	-------	---	---

60 >4335744 /41585

1356

	len =	2418	nex =	9	
	Term	10913	10690	-	0
5	Intr	11117	11009	-	0
	Intr	11446	11232	-	0
	Intr	11721	11575	-	0
	Intr	12110	11952	-	0
	Intr	12261	12202	-	0
10	Intr	12404	12337	-	0
	Intr	12578	12482	-	0
	Init	13107	12916	-	0
	>4335744	/35226			
15	len =	900	nex =	1	
	Sngl	39634	40533	+	0
20	>4335744	/41376			
	len =	2132	nex =	8	
	Init	47695	47718	+	0
25	Intr	47825	47884	+	0
	Intr	48035	48129	+	0
	Intr	48223	48395	+	0
	Intr	48485	48597	+	0
	Intr	48707	48752	+	0
30	Term	48859	48899	+	0
	Term	48974	49259	+	0
	>4335744	/1697			
35	len =	1601	nex =	3	
	Init	58565	58779	+	0
	Intr	59480	59534	+	0
	Term	59634	60165	+	0
40	>4337186	/33524			
	len =	173	nex =	1	
45	Sngl	2653	2825	+	0
	>4337186	/141837			
	len =	392	nex =	1	
50	Sngl	28822	28431	-	0
	>4337186	/110587			
55	len =	720	nex =	1	
	Sngl	39509	39644	+	0
	>4337186	/16295			
60					

1357

	len =	3190	nex =	11	
	Init	44064	44289	+	0
	Intr	44410	44537	+	0
5	Intr	44623	44694	+	0
	Intr	44771	44884	+	0
	Intr	45009	45077	+	0
	Intr	45167	45277	+	0
	Intr	45503	45565	+	0
10	Intr	45889	46044	+	0
	Intr	46155	46235	+	0
	Intr	46520	46648	+	0
	Term	46942	47253	+	0
15	>4337186	/7536			
	len =	2080	nex =	10	
	Term	62706	62383	-	0
20	Intr	62883	62815	-	0
	Intr	63047	62962	-	0
	Intr	63232	63146	-	0
	Intr	63400	63322	-	0
	Intr	63600	63523	-	0
25	Intr	63730	63689	-	0
	Intr	63936	63836	-	0
	Intr	64142	64032	-	0
	Init	64462	64286	-	0
30	>4337186	/6917			
	len =	277	nex =	1	
	Sngl	766	1042	+	0
35	>4371278	/15139			
	len =	1126	nex =	3	
	Init	15864	16132	+	0
	Intr	16428	16560	+	0
	Term	16673	16989	+	0
	>4371278	/23523			
45	len =	679	nex =	1	
	Sngl	18840	19518	+	0
50	>4371278	/108612			
	len =	457	nex =	1	
	Sngl	18849	19305	+	0
55	>4371278	/11998			
	len =	1361	nex =	4	
60	Term	20141	19887	-	0

1358

Intr	20371	20231	-	0
Intr	20758	20463	-	0
Init	21247	20880	-	0

5 >4371278 /25220

len = 1413 nex = 4

10	Init	24105	24468	+	0
	Intr	24590	24885	+	0
	Intr	24977	25117	+	0
	Term	25216	25517	+	0

>4371278 /38965

15 len = 1137 nex = 4

20	Init	24195	24468	+	0
	Intr	24590	24885	+	0
	Intr	24977	25117	+	0
	Term	25216	25331	+	0

>4371278 /3571

25 len = 1941 nex = 5

30	Init	38901	39030	+	0
	Intr	39208	39356	+	0
	Intr	39462	39650	+	0
	Intr	39739	39907	+	0
	Term	40334	40841	+	0

>4371278 /23542

35 len = 381 nex = 1

Sngl 4113 4343 + 0

>4371278 /1517

40 len = 850 nex = 2

	Init	43998	44170	+	0
45	Term	44506	44846	+	0

>4371278 /2454

len = 1371 nex = 3

50	Term	45469	45000	-	0
	Intr	45776	45574	-	0
	Init	46370	45982	-	0

>4371278 /38677

55 len = 1210 nex = 1

Sngl 46951 48155 + 0

60 >4376087 /92448

2025 RELEASE UNDER E.O. 14176

1359

	len =	708	nex =	2	
	Init	100575	100763	+	0
5	Term	100977	101282	+	0
>4376087 /30438					
	len =	2020	nex =	6	
10	Init	100575	100763	+	0
	Intr	100977	101158	+	0
	Intr	101420	101606	+	0
	Intr	101697	101858	+	0
15	Intr	101967	102269	+	0
	Term	102338	102594	+	0
>4376087 /18909					
20	len =	2031	nex =	9	
	Term	110792	110506	-	0
	Intr	110969	110872	-	0
	Intr	111157	111078	-	0
25	Intr	111325	111254	-	0
	Intr	111459	111418	-	0
	Intr	111712	111551	-	0
	Intr	112024	111796	-	0
	Intr	112222	112153	-	0
30	Init	112536	112404	-	0
>4376087 /26540					
	len =	1176	nex =	6	
35	Term	122365	122161	-	0
	Intr	122548	122446	-	0
	Intr	122806	122654	-	0
	Intr	122940	122913	-	0
40	Intr	123065	123029	-	0
	Init	123336	123143	-	0
>4376087 /18266					
45	len =	1184	nex =	2	
	Term	124539	123889	-	0
	Init	125072	124716	-	0
50	>4376087 /1794				
	len =	835	nex =	3	
	Term	12383	12097	-	0
55	Intr	12576	12475	-	0
	Init	12931	12658	-	0
>4376087 /11274					
60	len =	1278	nex =	5	

2025 RELEASE UNDER E.O. 14176

	>4388714	/39234		
5	len =	1951	nex =	3
	Init	79641	79729	+
	Intr	79806	79946	+
	Term	81346	81472	+
10	>4388714	/103540		
	len =	477	nex =	1
15	Sngl	82014	82484	+
	>4406752	/2021		
	len =	1990	nex =	5
20	Init	16077	16503	+
	Intr	16583	16721	+
	Intr	16810	16963	+
	Intr	17036	17159	+
	Term	17250	17624	+
25	>4406752	/20712		
	len =	1210	nex =	2
30	Init	42530	42826	+
	Term	42920	43154	+
	>4406752	/390		
35	len =	627	nex =	2
	Init	42528	42826	+
	Term	42920	43154	+
40	>4406752	/18785		
	len =	1039	nex =	2
	Term	43763	43148	-
45	Init	44186	43973	-
	>4406752	/38210		
	len =	1270	nex =	2
50	Term	64451	63872	-
	Init	65132	64859	-
	>4406752	/3964		
55	len =	1154	nex =	1
	Sngl	70202	69049	-
60	>4406752	/17089		

2025 RELEASE UNDER E.O. 14176

1362

	len =	2174	nex =	9	
	Init	77860	78033	+	0
5	Intr	78111	78291	+	0
	Intr	78423	78565	+	0
	Intr	78666	78729	+	0
	Intr	78889	78941	+	0
	Intr	79177	79232	+	0
10	Intr	79324	79398	+	0
	Intr	79482	79670	+	0
	Term	79755	80033	+	0
	>4406776	/36238			
15	len =	2338	nex =	4	
	Init	59059	59158	+	0
	Intr	59414	59731	+	0
20	Intr	59810	60366	+	0
	Term	60442	61396	+	0
	>4406776	/8529			
25	len =	528	nex =	1	
	Sngl	60863	61390	+	0
	>4406776	/39285			
30	len =	1992	nex =	0	
	>4406776	/33274			
35	len =	310	nex =	1	
	Sngl	61563	61868	+	0
	>4406790	/110175			
40	len =	473	nex =	1	
	Sngl	27736	27628	-	0
45	>4406790	/14605			
	len =	2614	nex =	9	
	Init	72541	72737	+	0
50	Intr	72828	72926	+	0
	Intr	73016	73126	+	0
	Intr	73401	73494	+	0
	Intr	73588	73688	+	0
	Intr	73852	73953	+	0
55	Intr	74050	74163	+	0
	Intr	74379	74453	+	0
	Term	74542	74819	+	0
	>4406790	/4489			
60					

2025-10-27 14:53:53

1363

	len =	396	nex =	1	
	Sngl	81642	81247	-	0
5	>4406805	/11178			
	len =	814	nex =	2	
10	Term	29635	29022	-	0
	Init	29835	29708	-	0
	>4406805	/112970			
15	len =	816	nex =	2	
	Term	29635	29024	-	0
	Init	29839	29708	-	0
	>4406805	/42057			
20	len =	156	nex =	1	
	Sngl	33661	33506	-	0
25	>4406805	/37664			
	len =	1390	nex =	2	
30	Term	32476	32343	-	0
	Init	33725	33309	-	0
	>4406805	/40712			
35	len =	1607	nex =	4	
	Term	34590	34119	-	0
	Intr	34839	34695	-	0
	Intr	35037	34925	-	0
	Init	35725	35374	-	0
40	>4406805	/35337			
	len =	1658	nex =	4	
45	Term	2559	2284	-	0
	Intr	2866	2649	-	0
	Intr	3066	2950	-	0
	Init	3279	3171	-	0
50	>4406805	/18197			
	len =	992	nex =	2	
	Term	38065	37812	-	0
55	Init	38803	38299	-	0
	>4406805	/105341			
60	len =	1033	nex =	1	

Reference No. 2750-942P

1364

	Sngl	51271	50239	-	0
	>4415905	/38827			
5	len =	1224	nex =	2	
	Init	11227	12103	+	0
	Term	12183	12450	+	0
10	>4415905	/14564			
	len =	1454	nex =	4	
	Term	2076	1784	-	0
15	Intr	2319	2171	-	0
	Intr	2509	2408	-	0
	Init	2694	2619	-	0
	>4415928	/11006			
20	len =	799	nex =	1	
	Sngl	15192	14394	-	0
25	>4415928	/10850			
	len =	819	nex =	2	
	Init	34311	34796	+	0
30	Term	34892	35129	+	0
	>4415928	/25575			
	len =	1297	nex =	3	
35	Term	35365	35116	-	0
	Intr	35883	35801	-	0
	Init	36412	36246	-	0
40	>4415928	/7802			
	len =	550	nex =	1	
	Sngl	60672	61212	+	0
45	>4415928	/33287			
	len =	394	nex =	1	
50	Sngl	60674	61067	+	0
	>4417264	/1841			
	len =	634	nex =	1	
55	Sngl	21959	22592	+	0
	>4417264	/38497			
60	len =	1733	nex =	2	

2025 RELEASE UNDER E.O. 14176

1365

	Term	50530	49193	-	0
	Init	50925	50620	-	0
5	>4417264 /31766				
	len =	2530	nex =	7	
	Term	51442	51135	-	0
10	Intr	51661	51539	-	0
	Intr	51937	51791	-	0
	Intr	52307	52197	-	0
	Intr	52643	52499	-	0
	Intr	53125	53032	-	0
15	Init	53657	53503	-	0
	>4417264 /122848				
	len =	1577	nex =	6	
20	Term	55897	55626	-	0
	Intr	56131	55980	-	0
	Intr	56288	56208	-	0
	Intr	56525	56379	-	0
25	Intr	56707	56608	-	0
	Init	57202	57038	-	0
	>4417264 /18922				
30	len =	2470	nex =	9	
	Init	82338	82391	+	0
	Intr	82475	82529	+	0
	Intr	82615	82687	+	0
35	Intr	82826	82913	+	0
	Intr	83016	83098	+	0
	Intr	83186	83315	+	0
	Intr	83414	83492	+	0
	Intr	83592	83656	+	0
40	Term	83736	84101	+	0
	>4417264 /33232				
	len =	2150	nex =	10	
45	Init	85855	86098	+	0
	Intr	86197	86292	+	0
	Intr	86405	86479	+	0
	Intr	86591	86659	+	0
50	Intr	86902	86978	+	0
	Intr	87066	87154	+	0
	Intr	87232	87361	+	0
	Intr	87439	87529	+	0
	Intr	87622	87687	+	0
55	Term	87762	88004	+	0
	>4417264 /10672				
60	len =	892	nex =	2	

2025-09-22 14:00:00

1366

	Term	88380	87971	-	0
	Init	88839	88459	-	0
	>4432811 /9846				
5	len =	1030	nex =	3	
	Init	12818	12984	+	0
	Intr	13233	13498	+	0
10	Term	13588	13843	+	0
	>4432829 /119200				
	len =	689	nex =	1	
15	Sngl	14377	13689	-	0
	>4432829 /14334				
20	len =	511	nex =	1	
	Sngl	21464	21974	+	0
	>4432829 /14111				
25	len =	2126	nex =	2	
	Term	37251	36808	-	0
	Init	37474	37342	-	0
30	>4432829 /10433				
	len =	935	nex =	2	
35	Term	50590	50253	-	0
	Init	51187	51056	-	0
	>4432829 /3673				
40	len =	833	nex =	1	
	Sngl	69750	70582	+	0
	>4432829 /23166				
45	len =	3474	nex =	7	
	Term	86292	85921	-	0
	Intr	86521	86378	-	0
50	Intr	87019	86949	-	0
	Intr	87222	87123	-	0
	Intr	87948	87918	-	0
	Intr	88192	88095	-	0
	Init	89394	89217	-	0
55	>4432829 /97776				
	len =	790	nex =	1	
60	Sngl	9122	8339	-	0

1367

	>4432847	/12895		
5	len =	2510	nex =	0
	>4432847	/10730		
	len =	676	nex =	2
10	Init	55664	55782	+
	Term	55970	56339	+
	>4432847	/97814		
15	len =	578	nex =	1
	Sngl	60009	60586	+
	>4454004	/32935		
20	len =	2261	nex =	7
	Term	17173	16861	-
	Intr	17285	17262	-
25	Intr	17408	17382	-
	Intr	17581	17514	-
	Intr	17828	17657	-
	Intr	18104	17916	-
	Init	19121	18868	-
30	>4454004	/29099		
	len =	2301	nex =	7
35	Term	17173	16834	-
	Intr	17285	17262	-
	Intr	17408	17382	-
	Intr	17581	17514	-
	Intr	17828	17657	-
40	Intr	18104	17916	-
	Init	18942	18868	-
	>4454004	/24398		
45	len =	2334	nex =	7
	Term	17173	16914	-
	Intr	17285	17262	-
	Intr	17408	17382	-
50	Intr	17581	17514	-
	Intr	17828	17657	-
	Intr	18104	17916	-
	Init	18942	18868	-
55	>4454004	/11559		
	len =	1123	nex =	4
	Init	2493	2663	+
60	Intr	2815	2904	+

2025 RELEASE UNDER E.O. 14176

1368

	Intr	2993	3115	+	0
	Term	3310	3615	+	0
	>4454004 /36536				
5	len =	1469	nex =	2	
	Term	31479	31022	-	0
	Init	32490	31900	-	0
10	>4454004 /29369				
	len =	1586	nex =	3	
15	Term	34339	33967	-	0
	Intr	34718	34509	-	0
	Init	35552	34914	-	0
	>4454004 /38603				
20	len =	1935	nex =	4	
	Init	42043	42575	+	0
	Intr	43023	43266	+	0
25	Intr	43356	43461	+	0
	Term	43541	43977	+	0
	>4454004 /26380				
30	len =	1279	nex =	3	
	Init	606	927	+	0
	Intr	1278	1392	+	0
	Term	1530	1881	+	0
35	>4454004 /150586				
	len =	714	nex =	1	
40	Sngl	66988	67701	+	0
	>4454004 /60				
	len =	1550	nex =	4	
45	Init	77637	77784	+	0
	Intr	78097	78344	+	0
	Intr	78424	78751	+	0
	Term	78826	79186	+	0
50	>4454004 /8252				
	len =	1648	nex =	1	
55	Sngl	9384	9995	+	0
	>4454022 /23878				
60	len =	1042	nex =	3	

1369

Term	11752	11125	-	0
Intr	11994	11828	-	0
Init	12166	12098	-	0

5 >4454022 /40423

len = 2478 nex = 10

Init	17803	17967	+	0
Intr	18546	18615	+	0
Intr	18803	18927	+	0
Intr	19076	19178	+	0
Intr	19253	19349	+	0
Intr	19432	19497	+	0
Intr	19591	19729	+	0
Intr	19823	19902	+	0
Intr	19987	20091	+	0
Term	20179	20280	+	0

20 >4454022 /6527

len = 1964 nex = 3

Term	33047	32545	-	0
Intr	33895	33766	-	0
Init	34508	34027	-	0

>4454022 /39185

30 len = 1852 nex = 5

Init	35769	36127	+	0
Intr	36534	36714	+	0
Intr	36797	36938	+	0
Intr	37035	37104	+	0
Term	37198	37620	+	0

>4454022 /14129

40 len = 1813 nex = 5

Init	35826	36127	+	0
Intr	36534	36714	+	0
Intr	36797	36938	+	0
Intr	37035	37104	+	0
Term	37198	37638	+	0

>4454022 /25397

50 len = 1077 nex = 4

Init	36551	36714	+	0
Intr	36797	36938	+	0
Intr	37035	37104	+	0
Term	37198	37627	+	0

>4454022 /37278

60 len = 2619 nex = 9

1370

	Term	42843	42630	-	0
	Intr	43042	42943	-	0
	Intr	43257	43130	-	0
	Intr	43493	43326	-	0
5	Intr	43710	43595	-	0
	Intr	43953	43801	-	0
	Intr	44188	44045	-	0
	Intr	44554	44268	-	0
	Init	45248	44799	-	0
10	>4454022 /37786				
	len =	310	nex =	1	
15	Sngl	50741	50436	-	0
	>4454022 /108949				
	len =	1140	nex =	3	
20	Term	50767	50437	-	0
	Intr	50907	50807	-	0
	Init	51576	51263	-	0
25	>4454022 /3802				
	len =	233	nex =	1	
	Sngl	55269	55037	-	0
30	>4454022 /41988				
	len =	4725	nex =	4	
35	Term	50767	50548	-	0
	Intr	50907	50807	-	0
	Intr	54477	51263	-	0
	Init	55272	55024	-	0
40	>4454022 /42101				
	len =	1286	nex =	2	
	Term	54477	53987	-	0
45	Init	55272	55024	-	0
	>4454022 /115402				
	len =	974	nex =	2	
50	Term	54477	54301	-	0
	Init	55272	55024	-	0
	>4454022 /118240				
55	len =	939	nex =	2	
	Term	54477	54334	-	0
	Init	55272	55024	-	0
60					

20250515 09:22:59

1371

>4454022 /32540

```
len =      938      nex =      2
```

5	Term	54477	54335	-	0
	Init	55272	55024	-	0

>4454022 /92204

```
10      len =      4728      nex =          4
```

Term	50767	50548	-	0
------	-------	-------	---	---

Intr	50907	50807	-	0
------	-------	-------	---	---

Intr	54477	51263	-	0
------	-------	-------	---	---

15	Init	55275	55024	-	0
----	------	-------	-------	---	---

>4454022 / 660

```
len = 730    nex = 1
```

Sngl	57533	56808	-	0
------	-------	-------	---	---

>4454022 /2946

```
25      len =      1489      nex =          3
```

Init	67882	68445	+	0
------	-------	-------	---	---

Intr	68793	68893	+	0
------	-------	-------	---	---

Term	69004	69370	+	0
------	-------	-------	---	---

30
 >4454022 /21949

```
len = 2013    nex = 7
```

```
35      Init      80581      80861      +      0
```

Intr	81338	81412	+	0
------	-------	-------	---	---

Intr	81487	81543	+	0
------	-------	-------	---	---

Intr	81634	81800	+	0
------	-------	-------	---	---

Intr	81901	82035	+	0
------	-------	-------	---	---

40	Intr	82128	82330	+	0
----	------	-------	-------	---	---

Term	82416	82593	+	0
------	-------	-------	---	---

>4454022 /122649

```
45      len =      1106      nex =          2
```

Init	99722	99910	+	0
------	-------	-------	---	---

Term	100566	100827	+	0
------	--------	--------	---	---

50 >4454447 /1898

```
len =      855    nex =      3
```

Init	105950	106053	+	0
------	--------	--------	---	---

55	Intr	106164	106220	+	0
----	------	--------	--------	---	---

Term	106331	106804	+	0
------	--------	--------	---	---

>4454447 /528

```
60      len =      1665      nex =      2
```

1372

	Term	107093	106729	-	0
	Init	108393	107558	-	0
	>4454447 /34908				
5	len =	2650	nex =	5	
	Term	109037	108541	-	0
	Intr	109327	109116	-	0
10	Intr	109565	109443	-	0
	Intr	109985	109788	-	0
	Init	111189	110312	-	0
	>4454447 /34827				
15	len =	1645	nex =	3	
	Init	118166	118441	+	0
	Intr	118556	118784	+	0
20	Term	119096	119810	+	0
	>4454447 /39669				
	len =	3441	nex =	11	
25	Term	14000	13685	-	0
	Intr	14573	14085	-	0
	Intr	14892	14646	-	0
	Intr	15056	14980	-	0
30	Intr	15442	15368	-	0
	Intr	15656	15555	-	0
	Intr	15876	15758	-	0
	Intr	16151	16028	-	0
	Intr	16297	16247	-	0
35	Intr	16489	16391	-	0
	Init	17125	16761	-	0
	>4454447 /16840				
40	len =	730	nex =	1	
	Sngl	26481	25758	-	0
	>4454447 /15103				
45	len =	670	nex =	1	
	Sngl	28997	29662	+	0
50	>4454447 /27566				
	len =	1108	nex =	5	
	Init	29987	30074	+	0
55	Intr	30173	30269	+	0
	Intr	30358	30436	+	0
	Intr	30525	30606	+	0
	Term	30777	31079	+	0
60	>4454447 /15863				

1373

	len =	3319	nex =	11	
	Init	38766	39223	+	0
5	Intr	39301	39429	+	0
	Intr	39523	39608	+	0
	Intr	39702	39828	+	0
	Intr	39928	39976	+	0
	Intr	40187	40269	+	0
10	Intr	40380	40447	+	0
	Intr	40895	40969	+	0
	Intr	41157	41244	+	0
	Intr	41368	41484	+	0
15	Term	41575	42084	+	0
	>4454447	/34875			
	len =	1750	nex =	5	
20	Term	42627	42287	-	0
	Intr	42873	42758	-	0
	Intr	43258	43042	-	0
	Intr	43617	43350	-	0
25	Init	44030	43729	-	0
	>4454447	/2512			
	len =	682	nex =	1	
30	Sngl	48986	49667	+	0
	>4454447	/41949			
	len =	430	nex =	1	
35	Sngl	48989	49412	+	0
	>4454447	/7215			
40	len =	566	nex =	1	
	Sngl	48989	49554	+	0
	>4454447	/120852			
45	len =	717	nex =	1	
	Sngl	48989	49705	+	0
50	>4454447	/10398			
	len =	1539	nex =	2	
	Init	61887	62281	+	0
55	Term	62817	63409	+	0
	>4454447	/91872			
60	len =	1055	nex =	2	

2025 RELEASE UNDER E.O. 14176

1374

Term	66889	66819	-	0
Init	67328	67140	-	0

>4454447 /14736

5	len =	970	nex =	2	
---	-------	-----	-------	---	--

Term	66889	66819	-	0
Init	67371	67140	-	0

10 >4454447 /33816

len =	389	nex =	1	
-------	-----	-------	---	--

15 Sngl 98602 98990 + 0

>4454585 /10680

20 len = 1458 nex = 4

Init	38285	38437	+	0
Intr	38805	38874	+	0
Intr	39093	39133	+	0
Term	39382	39742	+	0

25 >4454587 /105376

len =	670	nex =	1	
-------	-----	-------	---	--

30 Sngl 68863 69530 + 0

>4454587 /1268

35 len = 2668 nex = 11

Term	73570	73137	-	0
Intr	73845	73789	-	0
Intr	74008	73940	-	0
Intr	74245	74177	-	0
Intr	74415	74347	-	0
Intr	74602	74513	-	0
Intr	74986	74904	-	0
Intr	75192	75112	-	0
Intr	75362	75280	-	0
Intr	75529	75438	-	0
Init	75804	75648	-	0

>4454587 /159035

50 len = 1757 nex = 6

Term	76531	76459	-	0
Intr	76696	76662	-	0
Intr	76825	76768	-	0
Intr	76986	76926	-	0
Intr	77147	77068	-	0
Init	77788	77455	-	0

60 >4454587 /1045

1375

	len =	2613	nex =	6	
	Init	92251	92370	+	0
	Intr	92686	92822	+	0
5	Intr	93334	93583	+	0
	Intr	93678	93925	+	0
	Intr	94015	94422	+	0
	Term	94517	94863	+	0
10	>4455168 /4067				
	len =	1477	nex =	2	
	Init	39463	39668	+	0
15	Term	39795	40939	+	0
	>4455168 /279				
	len =	1488	nex =	2	
20	Init	39464	39668	+	0
	Term	40110	40951	+	0
	>4455168 /2059				
25	len =	2265	nex =	8	
	Term	41426	41124	-	0
	Intr	41759	41631	-	0
30	Intr	41920	41879	-	0
	Intr	42088	42029	-	0
	Intr	42279	42166	-	0
	Intr	42473	42383	-	0
	Intr	42767	42582	-	0
35	Init	43388	43192	-	0
	>4455168 /17912				
	len =	1491	nex =	2	
40	Init	48010	48378	+	0
	Term	49184	49500	+	0
	>4455168 /28717				
45	len =	1004	nex =	5	
	Term	65924	65732	-	0
	Intr	66098	66012	-	0
50	Intr	66357	66172	-	0
	Intr	66540	66446	-	0
	Init	66735	66635	-	0
	>4455168 /6091				
55	len =	1847	nex =	5	
	Init	71711	71850	+	0
	Intr	72177	72209	+	0
60	Intr	72291	72573	+	0

20250901 09:06:50

1376

	Intr	72646	73126	+	0
	Term	73255	73557	+	0
	>4455189 /37329				
5	len =	326	nex =	1	
	Sngl	77591	77916	+	0
10	>4455189 /34595				
	len =	1423	nex =	2	
	Init	16690	17210	+	0
15	Term	17612	17935	+	0
	>4455189 /31962				
	len =	1362	nex =	3	
20	Init	16690	16812	+	0
	Intr	17163	17210	+	0
	Term	17612	17944	+	0
25	>4455189 /111016				
	len =	1319	nex =	3	
	Init	16607	16812	+	0
30	Intr	17163	17210	+	0
	Term	17612	17925	+	0
	>4455189 /5367				
35	len =	1551	nex =	5	
	Init	184	470	+	0
	Intr	860	924	+	0
	Intr	1005	1077	+	0
40	Intr	1179	1232	+	0
	Term	1353	1734	+	0
	>4455189 /35284				
45	len =	1830	nex =	5	
	Term	33174	32856	-	0
	Intr	33398	33270	-	0
	Intr	33645	33478	-	0
50	Intr	34173	33730	-	0
	Init	34685	34427	-	0
	>4455189 /2574				
55	len =	1558	nex =	2	
	Init	35777	36240	+	0
	Term	36966	37334	+	0
60	>4455189 /33007				

2025 RELEASE UNDER E.O. 14176

1377

	len =	2392	nex =	0	
	>4455189	/29414			
5	len =	1398	nex =	3	
	Init	45338	45550	+	0
	Intr	46160	46309	+	0
10	Term	46391	46735	+	0
	>4455189	/11679			
	len =	1606	nex =	6	
15	Term	51378	51155	-	0
	Intr	51592	51511	-	0
	Intr	51844	51711	-	0
	Intr	52006	51919	-	0
20	Intr	52157	52111	-	0
	Init	52760	52482	-	0
	>4455189	/93281			
25	len =	1559	nex =	6	
	Term	51378	51202	-	0
	Intr	51592	51511	-	0
	Intr	51844	51711	-	0
30	Intr	52006	51919	-	0
	Intr	52157	52111	-	0
	Init	52760	52482	-	0
	>4455229	/17175			
35	len =	2816	nex =	7	
	Init	1117	1501	+	0
	Intr	1589	1775	+	0
40	Intr	1866	2116	+	0
	Intr	2667	2781	+	0
	Intr	2897	2989	+	0
	Intr	3429	3572	+	0
	Term	3664	3932	+	0
45	>4455229	/25830			
	len =	2596	nex =	10	
50	Init	16057	16292	+	0
	Intr	16817	16868	+	0
	Intr	16975	17170	+	0
	Intr	17248	17327	+	0
	Intr	17421	17537	+	0
55	Intr	17710	17760	+	0
	Intr	17847	17913	+	0
	Intr	18037	18140	+	0
	Intr	18220	18324	+	0
60	Term	18417	18652	+	0

2025 RELEASE UNDER E.O. 14176

1378

	>4455229	/28535		
	len =	1244	nex =	4
5	Init	20496	20665	+
	Intr	20786	20812	+
	Intr	20941	21247	+
	Term	21515	21739	+
10	>4455229	/111178		
	len =	1125	nex =	1
15	Sngl	42529	42245	-
	>4455229	/31461		
	len =	1664	nex =	3
20	Term	42587	42245	-
	Intr	43426	43256	-
	Init	43908	43679	-
25	>4455229	/3059		
	len =	1846	nex =	5
	Term	44919	44554	-
	Intr	45357	44992	-
30	Intr	46012	45582	-
	Intr	46216	46084	-
	Init	46399	46295	-
35	>4455229	/32974		
	len =	2037	nex =	6
	Term	47455	46948	-
	Intr	47942	47823	-
40	Intr	48307	48161	-
	Intr	48622	48409	-
	Intr	48744	48708	-
	Init	48984	48865	-
45	>4455229	/16463		
	len =	2152	nex =	4
	Init	57208	57329	+
50	Intr	57866	58020	+
	Intr	58096	58362	+
	Term	58556	59359	+
55	>4455262	/27914		
	len =	2620	nex =	11
	Init	101628	101891	+
	Intr	102198	102288	+
60	Intr	102390	102442	+

1379

	Intr	102552	102644	+	0
	Intr	102736	102807	+	0
	Intr	103074	103146	+	0
	Intr	103266	103311	+	0
5	Intr	103418	103481	+	0
	Intr	103589	103642	+	0
	Intr	103729	103806	+	0
	Term	103880	104247	+	0
10	>4455262 /4469				
	len =	1731	nex =	5	
	Init	104601	105127	+	0
15	Intr	105304	105424	+	0
	Intr	105524	105750	+	0
	Intr	105860	106150	+	0
	Term	106238	106331	+	0
20	>4455262 /6952				
	len =	2335	nex =	5	
	Init	105304	105424	+	0
25	Intr	105524	105750	+	0
	Intr	105860	106150	+	0
	Intr	106238	106443	+	0
	Term	106521	106935	+	0
30	>4455262 /17600				
	len =	670	nex =	1	
	Sngl	23292	23961	+	0
35	>4455262 /107817				
	len =	299	nex =	1	
40	Sngl	23641	23939	+	0
	>4455262 /11205				
	len =	595	nex =	1	
45	Sngl	76742	77336	+	0
	>4455262 /33780				
50	len =	670	nex =	1	
	Sngl	77699	78366	+	0
	>4455262 /11383				
55	len =	110	nex =	1	
	Sngl	92430	92539	+	0
60	>4455262 /18558				

1380

	len =	564	nex =	1	
5	Sngl	98939	99502	+	0
	>4455290	/33360			
	len =	1522	nex =	3	
10	Init	13247	13653	+	0
	Intr	14262	14502	+	0
	Term	14622	14768	+	0
	>4455290	/8361			
15	len =	1930	nex =	7	
	Term	34312	34003	-	0
	Intr	34498	34422	-	0
20	Intr	34712	34600	-	0
	Intr	35303	35236	-	0
	Intr	35416	35353	-	0
	Intr	35693	35629	-	0
	Init	35927	35782	-	0
25	>4455290	/15923			
	len =	2801	nex =	6	
30	Init	37303	37639	+	0
	Intr	37723	37815	+	0
	Intr	38133	38237	+	0
	Intr	38889	39041	+	0
	Intr	39597	39695	+	0
35	Term	39798	39926	+	0
	>4455290	/6626			
	len =	1582	nex =	5	
40	Term	4300	3883	-	0
	Intr	4575	4506	-	0
	Intr	4877	4654	-	0
	Intr	5062	4986	-	0
45	Init	5464	5337	-	0
	>4455290	/38327			
	len =	1909	nex =	6	
50	Term	4300	3745	-	0
	Intr	4575	4506	-	0
	Intr	4732	4654	-	0
	Intr	4877	4823	-	0
55	Intr	5062	4986	-	0
	Init	5653	5337	-	0
	>4455290	/8965			
60	len =	509	nex =	1	

1381

	Sngl	67245	66737	-	0
	>4455290 /40589				
5	len =	2272	nex =	8	
	Term	69364	68891	-	0
	Intr	69544	69452	-	0
10	Intr	69755	69618	-	0
	Intr	69969	69842	-	0
	Intr	70213	70053	-	0
	Intr	70386	70296	-	0
	Intr	70653	70597	-	0
15	Init	71162	70752	-	0
	>4455290 /7191				
	len =	1255	nex =	4	
20	Init	88336	88457	+	0
	Intr	88555	88623	+	0
	Intr	88706	88812	+	0
	Term	88909	89229	+	0
25	>4455290 /16314				
	len =	1256	nex =	4	
30	Init	88336	88457	+	0
	Intr	88555	88623	+	0
	Intr	88706	88812	+	0
	Term	88909	89230	+	0
35	>4455290 /124124				
	len =	464	nex =	2	
	Init	88791	88812	+	0
40	Term	88909	89254	+	0
	>4455321 /22274				
	len =	326	nex =	1	
45	Sngl	38721	39046	+	0
	>4455321 /4369				
50	len =	2293	nex =	7	
	Term	46985	46617	-	0
	Intr	47138	47080	-	0
	Intr	47302	47219	-	0
55	Intr	47493	47383	-	0
	Intr	47690	47591	-	0
	Intr	48106	47773	-	0
	Init	48909	48641	-	0
60	>4455321 /31030				

2025-09-24 14:29:00

1382

	len =	1700	nex =	6	
5	Term	53798	53289	-	0
	Intr	54066	53875	-	0
	Intr	54274	54158	-	0
	Intr	54493	54413	-	0
	Intr	54739	54643	-	0
	Init	54988	54836	-	0
10	>4455321 /32761				
	len =	2005	nex =	5	
15	Init	74537	74904	+	0
	Intr	75417	75734	+	0
	Intr	75871	75966	+	0
	Intr	76072	76139	+	0
	Term	76237	76541	+	0
20	>4455339 /15213				
	len =	1956	nex =	2	
25	Term	56170	55714	-	0
	Init	57669	56842	-	0
	>4455339 /36461				
30	len =	2000	nex =	5	
	Term	63860	63223	-	0
	Intr	64081	63950	-	0
	Intr	64602	64187	-	0
35	Intr	64869	64736	-	0
	Init	65222	64975	-	0
	>4455339 /33995				
40	len =	1663	nex =	7	
	Init	71457	71689	+	0
	Intr	71779	71876	+	0
	Intr	71985	72015	+	0
45	Intr	72096	72195	+	0
	Intr	72362	72432	+	0
	Intr	72557	72700	+	0
	Term	72814	73119	+	0
50	>4455348 /41557				
	len =	2280	nex =	4	
	Init	1059	1312	+	0
55	Intr	2117	2572	+	0
	Intr	2790	2930	+	0
	Term	3044	3338	+	0
60	>4455348 /32397				

1383

	len =	919	nex =	1	
	Sngl	35039	34121	-	0
5	>4455348	/31814			
	len =	1740	nex =	5	
10	Term	46233	45987	-	0
	Intr	46420	46319	-	0
	Intr	46639	46514	-	0
	Intr	46883	46799	-	0
	Init	47726	47605	-	0
15	>4455348	/11590			
	len =	1462	nex =	5	
20	Term	48273	48030	-	0
	Intr	48468	48348	-	0
	Intr	48653	48558	-	0
	Intr	48972	48817	-	0
	Init	49491	49206	-	0
25	>4455348	/18870			
	len =	1078	nex =	2	
30	Term	53197	52635	-	0
	Init	53712	53462	-	0
	>4455348	/39831			
35	len =	1473	nex =	2	
	Term	53197	52607	-	0
	Init	54079	53462	-	0
40	>4455348	/7873			
	len =	550	nex =	1	
	Sngl	70379	70404	+	0
45	>4455348	/42155			
	len =	1694	nex =	2	
50	Term	75967	75455	-	0
	Init	77148	76451	-	0
	>4467094	/32072			
55	len =	1215	nex =	2	
	Term	111621	110777	-	0
	Init	111991	111698	-	0
60	>4467094	/37966			

1384

	len =	1872	nex =	4	
	Init	113206	113830	+	0
	Intr	114252	114389	+	0
5	Intr	114536	114760	+	0
	Term	114833	115077	+	0
	>4467094 /39781				
10	len =	1570	nex =	3	
	Init	115216	115464	+	0
	Intr	116303	116503	+	0
	Term	116590	116779	+	0
15	>4467094 /38432				
	len =	1546	nex =	1	
20	Sngl	123453	124998	+	0
	>4467094 /36909				
	len =	715	nex =	2	
25	Term	18966	18835	-	0
	Init	19549	19051	-	0
	>4467094 /33799				
30	len =	3020	nex =	15	
	Term	20057	19801	-	0
	Intr	20251	20141	-	0
35	Intr	20444	20333	-	0
	Intr	20602	20525	-	0
	Intr	20750	20685	-	0
	Intr	20936	20832	-	0
	Intr	21221	21024	-	0
40	Intr	21392	21300	-	0
	Intr	21572	21480	-	0
	Intr	21781	21657	-	0
	Intr	21937	21861	-	0
	Intr	22163	22042	-	0
45	Intr	22345	22259	-	0
	Intr	22567	22517	-	0
	Init	22820	22674	-	0
	>4467094 /27692				
50	len =	678	nex =	2	
	Term	57759	57398	-	0
	Init	58075	57842	-	0
55	>4467094 /42757				
	len =	1316	nex =	2	
60	Term	61776	61514	-	0

20250316.091600

Reference No. 2750-942P

1385

	Init	62829	62037	-	0
	>4467094	/35382			
5	len =	654	nex =	1	
	Sngl	70975	71628	+	0
	>4467094	/18344			
10	len =	1002	nex =	4	
	Term	71933	71728	-	0
	Intr	72070	72017	-	0
15	Intr	72246	72141	-	0
	Init	72729	72335	-	0
	>4467094	/21			
20	len =	1308	nex =	5	
	Term	75122	74915	-	0
	Intr	75276	75205	-	0
	Intr	75455	75351	-	0
25	Intr	75807	75617	-	0
	Init	76222	76054	-	0
	>4467094	/7275			
30	len =	677	nex =	1	
	Sngl	89301	88625	-	0
	>4467094	/36035			
35	len =	1604	nex =	1	
	Sngl	90226	88623	-	0
40	>4467131	/14711			
	len =	954	nex =	5	
	Init	16249	16304	+	0
45	Intr	16399	16481	+	0
	Intr	16617	16701	+	0
	Intr	16800	16881	+	0
	Term	16973	17202	+	0
50	>4467131	/95661			
	len =	2264	nex =	0	
	>4467131	/34210			
55	len =	1513	nex =	4	
	Term	41365	41312	-	0
	Intr	41555	41457	-	0
60	Intr	41673	41645	-	0

2025 RELEASE UNDER E.O. 14176

[illegible]

Init	41919	41805	-	0
>4467131	/27810			
len =	2192	nex =	8	
Term	41365	41180	-	0
Intr	41555	41457	-	0
Intr	41673	41645	-	0
Intr	41919	41805	-	0
Intr	42078	42013	-	0
Intr	42823	42739	-	0
Intr	43052	42917	-	0
Init	43371	43235	-	0
>4467131	/10238			
len =	1433	nex =	2	
Init	44334	44556	+	0
Term	45193	45766	+	0
>4467131	/28204			
len =	1390	nex =	2	
Init	4668	5085	+	0
Term	5231	6052	+	0
>4467131	/6757			
len =	2370	nex =	10	
Term	67293	67054	-	0
Intr	67447	67379	-	0
Intr	67646	67546	-	0
Intr	67826	67730	-	0
Intr	68190	67911	-	0
Intr	68614	68402	-	0
Intr	68808	68694	-	0
Intr	69041	68936	-	0
Intr	69249	69136	-	0
Init	69423	69331	-	0
>4467131	/33640			
len =	2504	nex =	3	
Term	70027	69544	-	0
Intr	70463	70118	-	0
Init	72047	70911	-	0
>4467131	/18363			
len =	2290	nex =	7	
Init	73093	73341	+	0
Intr	73431	73471	+	0
Intr	73571	73730	+	0
Intr	73817	73992	+	0

1387

Intr	74069	74173	+	0
Intr	74253	74399	+	0
Term	74474	75115	+	0

5 >4467131 /32152

len = 1894 nex = 3

Term	76622	76090	-	0
Intr	77665	76996	-	0
Init	77983	77792	-	0

>4467131 /46

15 len = 1750 nex = 3

Term	79829	79698	-	0
Intr	80958	80259	-	0
Init	81238	81056	-	0

20 >4468103 /34690

len = 1595 nex = 4

Term	16660	16132	-	0
Intr	17189	16761	-	0
Intr	17559	17306	-	0
Init	17726	17639	-	0

30 >4468103 /11210

len = 2187 nex = 2

Term	42434	41751	-	0
Init	43937	42613	-	0

>4468103 /36616

40 len = 2338 nex = 9

Term	51987	51673	-	0
Intr	52425	52370	-	0
Intr	52655	52555	-	0
Intr	52938	52846	-	0
Intr	53108	53030	-	0
Intr	53260	53199	-	0
Intr	53449	53356	-	0
Intr	53625	53537	-	0
Init	54010	53706	-	0

50 >4468103 /8606

len = 2269 nex = 8

Term	60529	60215	-	0
Intr	60674	60616	-	0
Intr	60854	60790	-	0
Intr	61078	60942	-	0
Intr	61243	61191	-	0
Intr	61453	61376	-	0

2001-01-01 09:00:00

1388

	Intr	61736	61670	-	0
	Init	62483	62229	-	0
	>4468103 /943				
5	len =	2505	nex =	7	
	Init	73001	73391	+	0
	Intr	73776	73852	+	0
10	Intr	74018	74152	+	0
	Intr	74290	74344	+	0
	Intr	74437	74524	+	0
	Intr	74795	74893	+	0
	Term	74995	75505	+	0
15	>4468801 /41783				
	len =	574	nex =	1	
20	Sngl	1419	846	-	0
	>4468801 /40095				
	len =	2263	nex =	2	
25	Init	35213	36127	+	0
	Term	36724	37475	+	0
	>4468801 /33365				
30	len =	1750	nex =	2	
	Init	48345	49288	+	0
	Term	49374	50086	+	0
35	>4468801 /17990				
	len =	817	nex =	2	
40	Term	55584	55278	-	0
	Init	55732	55672	-	0
	>4468801 /121376				
45	len =	633	nex =	1	
	Sngl	80126	79494	-	0
	>4468801 /2583				
50	len =	1516	nex =	4	
	Init	82824	83081	+	0
	Intr	83164	83364	+	0
55	Intr	83438	83603	+	0
	Term	83816	84339	+	0
	>4468801 /33704				
60	len =	1125	nex =	2	

1389

	Term	91349	90695	-	0
	Init	91819	91488	-	0
5	>4468801	/23732			
	len =	2181	nex =	7	
	Init	94811	95064	+	0
10	Intr	95159	95396	+	0
	Intr	95517	95707	+	0
	Intr	95803	96038	+	0
	Intr	96147	96390	+	0
	Intr	96504	96611	+	0
15	Term	96769	96991	+	0
	>4468976	/92021			
	len =	1690	nex =	3	
20	Init	40307	40379	+	0
	Intr	40472	40537	+	0
	Term	40686	40979	+	0
25	>4468976	/41011			
	len =	640	nex =	1	
	Sngl	65865	66504	+	0
30	>4468976	/207350			
	len =	615	nex =	1	
35	Sngl	88477	89091	+	0
	>4469002	/39733			
	len =	1604	nex =	1	
40	Sngl	11302	9699	-	0
	>4469002	/7756			
45	len =	731	nex =	2	
	Term	15799	15454	-	0
	Init	16184	15960	-	0
50	>4469002	/33380			
	len =	1150	nex =	2	
	Term	1303	530	-	0
55	Init	1678	1406	-	0
	>4469002	/20829			
60	len =	3120	nex =	9	

1390

	Term	17041	16572	-	0
	Intr	17196	17137	-	0
	Intr	17550	17459	-	0
	Intr	17871	17773	-	0
5	Intr	18146	18054	-	0
	Intr	18282	18225	-	0
	Intr	18836	18687	-	0
	Intr	19200	18957	-	0
	Init	19691	19451	-	0
10	>4469002 /11048				
	len =	701	nex =	2	
15	Init	20958	21120	+	0
	Term	21217	21658	+	0
	>4469002 /799				
20	len =	2411	nex =	5	
	Term	32961	32391	-	0
	Intr	33385	33177	-	0
	Intr	33762	33645	-	0
25	Intr	34243	33981	-	0
	Init	34801	34389	-	0
	>4469002 /115976				
30	len =	1533	nex =	4	
	Term	3286	2986	-	0
	Intr	3441	3370	-	0
	Intr	3876	3740	-	0
35	Init	4518	3982	-	0
	>4469002 /8827				
	len =	490	nex =	1	
40	Sngl	61418	61899	+	0
	>4469002 /12935				
45	len =	473	nex =	1	
	Sngl	62817	63289	+	0
	>4469002 /14002				
50	len =	1424	nex =	6	
	Term	76434	76173	-	0
	Intr	76611	76509	-	0
55	Intr	76843	76705	-	0
	Intr	77038	76946	-	0
	Intr	77166	77120	-	0
	Init	77596	77275	-	0
60	>4469002 /24162				

2025-06-10 09:10:00

1391

	len =	1297	nex =	3	
	Init	79053	79213	+	0
5	Intr	79319	79691	+	0
	Term	79762	80345	+	0
	>4469002 /107201				
10	len =	1003	nex =	2	
	Term	84417	83587	-	0
	Init	84589	84496	-	0
15	>4490291 /30445				
	len =	683	nex =	2	
	Init	30362	30506	+	0
20	Term	30547	30757	+	0
	>4490291 /37550				
	len =	678	nex =	1	
25	Sngl	30367	31044	+	0
	>4490291 /107617				
30	len =	730	nex =	3	
	Init	35200	35539	+	0
	Intr	35630	35718	+	0
	Term	35791	35920	+	0
35	>4490291 /119671				
	len =	1770	nex =	8	
40	Term	2571	2225	-	0
	Intr	2837	2755	-	0
	Intr	3098	3032	-	0
	Intr	3230	3180	-	0
	Intr	3408	3296	-	0
45	Intr	3574	3461	-	0
	Intr	3724	3668	-	0
	Init	3994	3815	-	0
	>4490291 /26806				
50	len =	1786	nex =	7	
	Term	2571	2225	-	0
	Intr	2837	2755	-	0
55	Intr	3098	3032	-	0
	Intr	3230	3180	-	0
	Intr	3574	3461	-	0
	Intr	3724	3668	-	0
	Init	3995	3821	-	0
60					

2025-09-22 09:56:00

1392

>4490291 /32346

len = 1453 nex = 4

5	Init	48142	48411	+	0
	Intr	48498	48686	+	0
	Intr	48784	48946	+	0
	Term	49036	49594	+	0

10 >4490291 /19165

len = 1810 nex = 5

	Term	50287	49800	-	0
15	Intr	50763	50369	-	0
	Intr	51202	50861	-	0
	Intr	51411	51280	-	0
	Init	51600	51524	-	0

20 >4490291 /40596

len = 2940 nex = 9

	Term	69161	68951	-	0
25	Intr	69309	69244	-	0
	Intr	69681	69649	-	0
	Intr	69887	69828	-	0
	Intr	70245	70184	-	0
	Intr	70842	70771	-	0
30	Intr	71090	71045	-	0
	Intr	71267	71170	-	0
	Init	71890	71358	-	0

>4490291 /24738

len = 2146 nex = 8

	Term	72209	71984	-	0
	Intr	72366	72294	-	0
40	Intr	72656	72456	-	0
	Intr	72877	72736	-	0
	Intr	73193	72966	-	0
	Intr	73516	73293	-	0
	Intr	73720	73603	-	0
45	Init	74129	73817	-	0

>4490291 /37432

len = 2266 nex = 5

50	Init	79757	80261	+	0
	Intr	80346	80623	+	0
	Intr	80743	81019	+	0
	Intr	81256	81501	+	0
55	Term	81593	82022	+	0

>4490324 /39641

len = 221 nex = 1

60

1393

	Sngl	18671	18451	-	0
	>4490324	/150250			
5	len =	264	nex =	1	
	Sngl	18806	18543	-	0
	>4490324	/94851			
10	len =	264	nex =	1	
	Sngl	30283	30027	-	0
15	>4490324	/41363			
	len =	709	nex =	1	
	Sngl	30771	30063	-	0
20	>4490324	/119544			
	len =	454	nex =	1	
25	Sngl	43886	43433	-	0
	>4490324	/2315			
	len =	1640	nex =	2	
30	Term	44776	43494	-	0
	Init	45133	44870	-	0
	>4490324	/21759			
35	len =	1815	nex =	6	
	Init	57939	58220	+	0
	Intr	58586	58785	+	0
40	Intr	58863	59003	+	0
	Intr	59082	59162	+	0
	Intr	59290	59414	+	0
	Term	59498	59753	+	0
45	>4490324	/37986			
	len =	1820	nex =	6	
	Init	57939	58220	+	0
50	Intr	58586	58785	+	0
	Intr	58863	59003	+	0
	Intr	59082	59162	+	0
	Intr	59290	59414	+	0
	Term	59498	59758	+	0
55	>4490324	/17181			
	len =	812	nex =	1	
60	Sngl	61230	61139	-	0

1394

>4490324 /10140

5	len =	550	nex =	1	
	Sngl	72094	71548	-	0

>4490324 /14973

10	len =	508	nex =	1	
	Sngl	72740	73247	+	0

>4490324 /117474

15	len =	1555	nex =	5	
	Term	82576	82369	-	0
	Intr	82830	82661	-	0
20	Intr	83046	82926	-	0
	Intr	83406	83125	-	0
	Init	83798	83526	-	0

>4490701 /6867

25	len =	1007	nex =	4	
	Term	14504	14298	-	0
	Intr	14711	14599	-	0
30	Intr	14909	14780	-	0
	Init	15294	15202	-	0

>4490701 /4374

35	len =	986	nex =	4	
	Term	14504	14319	-	0
	Intr	14711	14599	-	0
	Intr	14909	14780	-	0
40	Init	15294	15202	-	0

>4490701 /41420

45	len =	1297	nex =	5	
	Term	14504	14300	-	0
	Intr	14711	14599	-	0
	Intr	14909	14780	-	0
	Intr	15303	15202	-	0
50	Init	15596	15529	-	0

>4490701 /39825

55	len =	2359	nex =	8	
	Term	22765	22319	-	0
	Intr	22967	22931	-	0
	Intr	23112	23056	-	0
	Intr	23485	23327	-	0
60	Intr	23674	23573	-	0

2025 RELEASE UNDER E.O. 14176

1395

Intr	23973	23864	-	0
Intr	24351	24252	-	0
Init	24677	24452	-	0

5 >4490701 /21752

len = 2350 nex = 7

10	Term	22765	22390	-	0
	Intr	22967	22931	-	0
	Intr	23112	23056	-	0
	Intr	23485	23327	-	0
	Intr	23674	23573	-	0
	Intr	23973	23864	-	0
15	Init	24351	24252	-	0

>4490701 /5206

20 len = 1848 nex = 5

	Term	25202	24945	-	0
	Intr	25336	25283	-	0
	Intr	25666	25461	-	0
25	Intr	25913	25833	-	0
	Init	26792	26587	-	0

>4490717 /8812

30 len = 1259 nex = 5

	Term	101847	101489	-	0
	Intr	102100	101932	-	0
	Intr	102370	102201	-	0
	Intr	102632	102454	-	0
35	Init	102744	102713	-	0

>4490717 /35272

40 len = 3745 nex = 12

	Term	101847	101520	-	0
	Intr	102100	101932	-	0
	Intr	102370	102201	-	0
	Intr	102632	102454	-	0
45	Intr	102778	102713	-	0
	Intr	103035	102862	-	0
	Intr	103512	103214	-	0
	Intr	103702	103602	-	0
	Intr	103878	103794	-	0
50	Intr	104125	103972	-	0
	Intr	104415	104219	-	0
	Init	105069	104785	-	0

>4490717 /36691

55 len = 1469 nex = 3

	Init	66866	67420	+	0
	Intr	67516	67595	+	0
60	Term	67696	68334	+	0

2025 RELEASE UNDER E.O. 14176

1396

>4490717 /142634

5	len =	1930	nex =	4	
	Term	74823	74205	-	0
	Intr	75637	75441	-	0
	Intr	75819	75719	-	0
	Init	76126	75910	-	0

10 >4490717 /8286

	len =	1301	nex =	4	
15	Term	86445	86257	-	0
	Intr	86683	86641	-	0
	Intr	87034	86928	-	0
	Init	87557	87378	-	0

20 >4490734 /16697

	len =	1570	nex =	5	
	Term	99947	99849	-	0
25	Intr	100104	100038	-	0
	Intr	100447	100205	-	0
	Intr	100749	100543	-	0
	Init	101411	100828	-	0

30 >4490734 /15747

	len =	1796	nex =	5	
	Init	37430	37611	+	0
35	Intr	37696	37769	+	0
	Intr	38182	38305	+	0
	Intr	38387	38698	+	0
	Term	38786	39225	+	0

40 >4490734 /20457

	len =	761	nex =	1	
	Sngl	42123	42883	+	0

45 >4490734 /123060

	len =	2073	nex =	7	
50	Term	46245	45961	-	0
	Intr	46442	46334	-	0
	Intr	46578	46536	-	0
	Intr	46988	46695	-	0
	Intr	47215	47145	-	0
55	Intr	47426	47318	-	0
	Init	48033	47855	-	0

>4490734 /37341

60	len =	1908	nex =	10	
----	-------	------	-------	----	--

2025 RELEASE UNDER E.O. 14176

1397

	Term	74312	74162	-	0
	Intr	74472	74426	-	0
	Intr	74624	74565	-	0
5	Intr	74765	74706	-	0
	Intr	74914	74859	-	0
	Intr	75047	75014	-	0
	Intr	75206	75132	-	0
	Intr	75344	75294	-	0
10	Intr	75462	75418	-	0
	Init	76069	75542	-	0
>4490734		/24741			
15	len =	2473	nex =	5	
	Init	77167	77764	+	0
	Intr	77835	78112	+	0
	Intr	78224	78500	+	0
20	Intr	78836	79081	+	0
	Term	79167	79639	+	0
>4490734		/19562			
25	len =	1739	nex =	4	
	Init	77929	78112	+	0
	Intr	78224	78500	+	0
	Intr	78836	79081	+	0
30	Term	79167	79667	+	0
>4490734		/8545			
35	len =	1258	nex =	2	
	Init	80276	80540	+	0
	Term	80921	81533	+	0
>4490734		/121755			
40	len =	936	nex =	1	
	Sngl	91697	90762	-	0
45	>4490734	/9448			
	len =	1168	nex =	3	
	Term	93687	93247	-	0
50	Intr	93849	93773	-	0
	Init	94003	93932	-	0
>4490734		/4532			
55	len =	1270	nex =	3	
	Term	93687	93204	-	0
	Intr	93849	93773	-	0
	Init	94003	93932	-	0
60					

1398

>4490734 /36190

len = 1275 nex = 4

5	Term	93687	93284	-	0
	Intr	93849	93773	-	0
	Intr	94003	93932	-	0
	Init	94558	94467	-	0

10 >4490734 /2670

len = 1412 nex = 4

15	Term	95295	94885	-	0
	Intr	95484	95394	-	0
	Intr	95663	95587	-	0
	Init	95822	95751	-	0

20 >4490734 /8313

len = 741 nex = 1

Sngl 96873 97152 + 0

25 >4490734 /22456

len = 802 nex = 2

30	Term	97920	97406	-	0
	Init	98207	98150	-	0

>4510323 /29630

len = 1699 nex = 8

35	Init	116587	116710	+	0
	Intr	116800	116892	+	0
	Intr	117017	117121	+	0
	Intr	117204	117303	+	0
40	Intr	117393	117484	+	0
	Intr	117588	117668	+	0
	Intr	117770	117838	+	0
	Term	118019	118285	+	0

45 >4510323 /30217

len = 1650 nex = 8

50	Init	116589	116710	+	0
	Intr	116800	116892	+	0
	Intr	117017	117121	+	0
	Intr	117204	117303	+	0
	Intr	117393	117484	+	0
	Intr	117588	117668	+	0
55	Intr	117770	117838	+	0
	Term	118019	118238	+	0

>4510323 /7542

60 len = 1518 nex = 4

1399

	Term	20387	20042	-	0
	Intr	20566	20495	-	0
	Intr	20825	20709	-	0
5	Init	21559	20906	-	0
>4510323 /36518					
	len =	2110	nex =	0	
10	>4510323 /16583				
	len =	370	nex =	1	
15	Sngl	53962	53601	-	0
>4510323 /11468					
	len =	1887	nex =	7	
20	Init	56611	56922	+	0
	Intr	57316	57415	+	0
	Intr	57509	57576	+	0
	Intr	57795	57847	+	0
25	Intr	57934	57982	+	0
	Intr	58057	58104	+	0
	Term	58233	58497	+	0
>4510323 /28726					
30	len =	1150	nex =	3	
	Term	64164	63855	-	0
	Intr	64685	64611	-	0
35	Init	65000	64851	-	0
>4510323 /4776					
	len =	1302	nex =	5	
40	Init	84573	84896	+	0
	Intr	84980	85098	+	0
	Intr	85207	85395	+	0
	Intr	85480	85531	+	0
45	Term	85613	85874	+	0
>4510323 /21074					
	len =	1418	nex =	3	
50	Term	85969	85508	-	0
	Intr	86191	86055	-	0
	Init	86925	86535	-	0
55	>4510323 /13256				
	len =	1850	nex =	3	
	Term	94960	94737	-	0
60	Intr	95223	95079	-	0

1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2

0

	Init	96586	95450	-	0
	>4510338	/96414			
5	len =	1937	nex =	3	
	Term	28550	28235	-	0
	Intr	29151	29012	-	0
10	Init	30171	29548	-	0
	>4510338	/24255			
	len =	971	nex =	1	
15	Sngl	34528	33558	-	0
	>4510338	/2576			
	len =	2237	nex =	9	
20	Init	36933	37151	+	0
	Intr	37427	37542	+	0
	Intr	37638	37697	+	0
25	Intr	37834	38016	+	0
	Intr	38097	38201	+	0
	Intr	38291	38387	+	0
	Intr	38479	38586	+	0
	Intr	38668	38803	+	0
30	Term	38888	39169	+	0
	>4510338	/156731			
	len =	1390	nex =	2	
35	Init	44273	44363	+	0
	Term	45034	45662	+	0
	>4510338	/114613			
40	len =	1210	nex =	0	
	>4510338	/5167			
	len =	2170	nex =	5	
45	Init	52854	52965	+	0
	Intr	53527	53789	+	0
	Intr	53878	54089	+	0
	Intr	54172	54439	+	0
50	Term	54516	55019	+	0
	>4510338	/3668			
	len =	1690	nex =	2	
55	Init	56164	56327	+	0
	Term	57130	57851	+	0
	>4510338	/2813			
60					

1401

	len =	1832	nex =	7	
	Init	68165	68340	+	0
	Intr	68439	68646	+	0
5	Intr	68750	68835	+	0
	Intr	68919	69020	+	0
	Intr	69117	69250	+	0
	Intr	69373	69421	+	0
	Term	69637	69988	+	0
10	>4510338 /12251				
	len =	1059	nex =	1	
15	Sngl	71208	70150	-	0
	>4510360 /39036				
	len =	2077	nex =	8	
20	Term	101338	101020	-	0
	Intr	101452	101406	-	0
	Intr	101719	101533	-	0
	Intr	102088	101798	-	0
25	Intr	102369	102175	-	0
	Intr	102562	102461	-	0
	Intr	102944	102645	-	0
	Init	103096	103017	-	0
30	>4510360 /36041				
	len =	1536	nex =	5	
	Term	102088	101812	-	0
35	Intr	102369	102175	-	0
	Intr	102562	102461	-	0
	Intr	102944	102645	-	0
	Init	103095	103017	-	0
40	>4510360 /21707				
	len =	2395	nex =	7	
	Term	101338	101025	-	0
45	Intr	101452	101406	-	0
	Intr	101719	101533	-	0
	Intr	102088	101798	-	0
	Intr	102369	102175	-	0
	Intr	102562	102461	-	0
50	Init	102944	102645	-	0
	>4510360 /28577				
	len =	1009	nex =	3	
55	Term	103942	103748	-	0
	Intr	104361	104019	-	0
	Init	104756	104575	-	0
60	>4510360 /22535				

1402

	len =	1330	nex =	3	
	Term	105113	104790	-	0
5	Intr	105546	105210	-	0
	Init	106116	105944	-	0
	>4510360	/39503			
10	len =	1090	nex =	3	
	Term	107026	106783	-	0
	Intr	107457	107118	-	0
	Init	107872	107683	-	0
15	>4510360	/29951			
	len =	1349	nex =	4	
20	Init	12868	13060	+	0
	Intr	13411	13550	+	0
	Intr	13677	13802	+	0
	Term	13876	14216	+	0
25	>4510360	/13310			
	len =	2350	nex =	8	
	Init	31440	31565	+	0
30	Intr	31864	31919	+	0
	Intr	32001	32107	+	0
	Intr	32534	32590	+	0
	Intr	32698	32761	+	0
	Intr	32865	32999	+	0
35	Intr	33109	33192	+	0
	Term	33299	33781	+	0
	>4510360	/26418			
40	len =	874	nex =	1	
	Sngl	46437	45564	-	0
	>4510360	/19481			
45	len =	806	nex =	1	
	Sngl	49946	49141	-	0
50	>4510360	/91908			
	len =	732	nex =	1	
	Sngl	5525	6256	+	0
55	>4510392	/16423			
	len =	1588	nex =	2	
60	Term	35088	34032	-	0

1404

	Intr	46949	47090	+	0
	Intr	47177	47246	+	0
	Intr	47334	47399	+	0
	Intr	47531	47653	+	0
5	Term	47737	48053	+	0
	>4512646 /41045				
	len =	1632	nex =	2	
10	Term	48617	48420	-	0
	Init	50051	49809	-	0
	>4512646 /40190				
15	len =	1430	nex =	2	
	Term	49533	49412	-	0
	Init	50053	49809	-	0
20	>4512646 /96543				
	len =	1630	nex =	3	
25	Term	48617	48428	-	0
	Intr	49533	49412	-	0
	Init	50057	49809	-	0
	>4512646 /15689				
30	len =	1669	nex =	3	
	Term	48617	48425	-	0
	Intr	49533	49412	-	0
35	Init	50093	49809	-	0
	>4512646 /41317				
40	len =	2050	nex =	8	
	Term	50564	50329	-	0
	Intr	50835	50657	-	0
	Intr	51046	50938	-	0
	Intr	51222	51137	-	0
45	Intr	51577	51518	-	0
	Intr	51893	51678	-	0
	Intr	52086	51973	-	0
	Init	52376	52179	-	0
50	>4512656 /41712				
	len =	1391	nex =	4	
	Term	105749	105453	-	0
55	Intr	105920	105838	-	0
	Intr	106277	106194	-	0
	Init	106843	106633	-	0
60	>4512656 /125409				

2025-09-23 14:04

1405

	len =	971	nex =	1	
	Sngl	14193	15163	+	0
5	>4512656	/35051			
	len =	2301	nex =	7	
10	Term	23482	23178	-	0
	Intr	23695	23596	-	0
	Intr	24031	23903	-	0
	Intr	24198	24130	-	0
	Intr	24399	24304	-	0
	Intr	24814	24499	-	0
15	Init	25478	25178	-	0
	>4512656	/102453			
20	len =	581	nex =	1	
	Sngl	44496	45076	+	0
	>4512656	/42863			
25	len =	1078	nex =	2	
	Term	47194	46904	-	0
	Init	47981	47309	-	0
30	>4512656	/41214			
	len =	1304	nex =	3	
	Init	54263	54446	+	0
35	Intr	54531	54981	+	0
	Term	55077	55566	+	0
	>4512656	/41345			
40	len =	1301	nex =	2	
	Init	54266	54981	+	0
	Term	55077	55566	+	0
45	>4512656	/99303			
	len =	1363	nex =	3	
	Init	54266	54446	+	0
50	Intr	54876	54981	+	0
	Term	55077	55628	+	0
	>4512656	/156913			
55	len =	550	nex =	1	
	Sngl	54268	54446	+	0
60	>4512656	/157512			

1406

	len =	265	nex =	1	
	Sngl	55360	55624	+	0
5	>4512656	/35604			
	len =	2847	nex =	3	
10	Term	58659	58568	-	0
	Intr	58861	58780	-	0
	Init	61159	60918	-	0
	>4512656	/17416			
15	len =	3176	nex =	13	
	Init	75697	75985	+	0
	Intr	76075	76255	+	0
20	Intr	76358	76443	+	0
	Intr	76651	76755	+	0
	Intr	76872	76934	+	0
	Intr	77055	77164	+	0
	Intr	77316	77435	+	0
25	Intr	77507	77582	+	0
	Intr	77677	77750	+	0
	Intr	77840	77960	+	0
	Intr	78076	78173	+	0
	Intr	78255	78378	+	0
30	Term	78474	78872	+	0
	>4512656	/21855			
	len =	1706	nex =	5	
35	Init	81740	82100	+	0
	Intr	82425	82580	+	0
	Intr	82665	82797	+	0
	Intr	82880	82919	+	0
40	Term	83024	83445	+	0
	>4512656	/2928			
	len =	1721	nex =	5	
45	Init	81740	82100	+	0
	Intr	82425	82580	+	0
	Intr	82665	82797	+	0
	Intr	82880	82919	+	0
50	Term	83024	83460	+	0
	>4512656	/26194			
	len =	610	nex =	1	
55	Sngl	84529	85135	+	0
	>4512656	/17723			
	len =	1334	nex =	5	
60					

1407

	Term	88998	88678	-	0
	Intr	89288	89093	-	0
	Intr	89677	89635	-	0
	Intr	89871	89764	-	0
5	Init	90011	89950	-	0
	>4512656 /2227				
10	len =	1692	nex =	3	
	Term	98186	98145	-	0
	Intr	98404	98327	-	0
	Init	99420	98801	-	0
15	>4512690 /91902				
	len =	1930	nex =	4	
20	Init	13553	13670	+	0
	Intr	13990	14118	+	0
	Intr	14808	14995	+	0
	Term	15104	15473	+	0
25	>4512690 /20429				
	len =	3383	nex =	11	
	Term	22804	22366	-	0
	Intr	23013	22893	-	0
30	Intr	23189	23121	-	0
	Intr	23389	23266	-	0
	Intr	23599	23484	-	0
	Intr	24010	23879	-	0
	Intr	24292	24222	-	0
35	Intr	24490	24406	-	0
	Intr	24696	24611	-	0
	Intr	24878	24814	-	0
	Init	25461	25365	-	0
40	>4512690 /36904				
	len =	2530	nex =	11	
	Term	4054	3778	-	0
45	Intr	4266	4144	-	0
	Intr	4517	4353	-	0
	Intr	4741	4607	-	0
	Intr	4977	4825	-	0
	Intr	5158	5084	-	0
50	Intr	5345	5268	-	0
	Intr	5505	5441	-	0
	Intr	5714	5603	-	0
	Intr	5907	5797	-	0
	Init	6307	6091	-	0
55	>4512690 /30611				
	len =	1786	nex =	5	
60	Init	72103	72432	+	0

2025-03-26 09:16:00

1408

	Intr	72543	72616	+	0
	Intr	73261	73335	+	0
	Intr	73495	73521	+	0
5	Term	73686	73888	+	0
	>4512690 /27629				
	len =	970	nex =	3	
10	Term	74223	74064	-	0
	Intr	74366	74301	-	0
	Init	75026	74717	-	0
	>4512690 /6967				
15	len =	2086	nex =	5	
	Term	84416	84137	-	0
	Intr	84712	84524	-	0
20	Intr	84919	84815	-	0
	Intr	85308	85275	-	0
	Init	86222	85706	-	0
	>4512690 /13277				
25	len =	2026	nex =	7	
	Term	7610	7394	-	0
	Intr	8243	7704	-	0
30	Intr	8397	8338	-	0
	Intr	8529	8479	-	0
	Intr	8713	8646	-	0
	Intr	8966	8823	-	0
	Init	9212	9088	-	0
35	>4519183 /205761				
	len =	1035	nex =	2	
40	Init	26007	26147	+	0
	Term	26837	27041	+	0
	>4519183 /3325				
45	len =	1098	nex =	3	
	Init	31470	31599	+	0
	Intr	31784	31867	+	0
	Term	32242	32567	+	0
50	>4519183 /25308				
	len =	1078	nex =	3	
55	Init	31473	31599	+	0
	Intr	31784	31867	+	0
	Term	32242	32550	+	0
	>4519183 /111672				
60					

2025-09-26 16:00

1410

	Intr	32336	32265	-	0
	Init	32505	32367	-	0
5	>4519187	/104871			
	len =	350	nex =	1	
	Sngl	32958	32609	-	0
10	>4519187	/2924			
	len =	1734	nex =	3	
15	Term	31639	31291	-	0
	Intr	32336	32265	-	0
	Init	33024	32367	-	0
	>4519187	/29696			
20	len =	1707	nex =	7	
	Init	48852	48942	+	0
	Intr	49200	49468	+	0
25	Intr	49539	49643	+	0
	Intr	49730	49816	+	0
	Intr	49902	49960	+	0
	Intr	50084	50237	+	0
	Term	50302	50558	+	0
30	>4519187	/25264			
	len =	2718	nex =	8	
	Init	53157	53318	+	0
35	Intr	53816	53957	+	0
	Intr	54090	54186	+	0
	Intr	54263	54332	+	0
	Intr	54428	54532	+	0
	Intr	54613	54711	+	0
40	Intr	55041	55128	+	0
	Term	55241	55522	+	0
	>4519188	/17098			
45	len =	850	nex =	2	
	Term	14912	14578	-	0
	Init	15424	14997	-	0
50	>4519188	/95892			
	len =	1436	nex =	6	
	Init	29371	29651	+	0
55	Intr	29739	29792	+	0
	Intr	29890	29970	+	0
	Intr	30055	30124	+	0
	Intr	30331	30419	+	0
	Term	30560	30806	+	0
60					

1411

>4519188 /251

len = 1462 nex = 6

5	Init	29379	29651	+	0
	Intr	29739	29792	+	0
	Intr	29890	29970	+	0
	Intr	30055	30124	+	0
	Intr	30331	30419	+	0
10	Term	30560	30840	+	0

>4519190 /158397

len = 1101 nex = 2

15	Init	31318	31483	+	0
	Term	31775	32418	+	0

>4519190 /15702

len = 3270 nex = 7

20	Init	37427	38279	+	0
	Intr	38619	38903	+	0
25	Intr	38977	39195	+	0
	Intr	39285	39464	+	0
	Intr	39540	40031	+	0
	Intr	40123	40284	+	0
30	Term	40365	40696	+	0

>4519191 /26549

len = 734 nex = 2

35	Init	35516	35748	+	0
	Term	35845	36249	+	0

>4519191 /35233

len = 1554 nex = 2

40	Term	37318	36349	-	0
	Init	37902	37399	-	0

45 >4519191 /5748

len = 730 nex = 2

50	Term	38416	37997	-	0
	Init	38725	38497	-	0

>4519191 /97675

len = 1914 nex = 4

55	Init	45399	45522	+	0
	Intr	45674	45748	+	0
	Intr	46594	46904	+	0
60	Term	47000	47312	+	0

1412

>4519191 /17006

len = 2118 nex = 3

5	Init	47579	48385	+	0
	Intr	48874	49077	+	0
	Term	49162	49696	+	0

>4519191 /6244

10 len = 3780 nex = 13

	Init	59272	59472	+	0
	Intr	59565	59603	+	0
15	Intr	59866	60033	+	0
	Intr	60122	60227	+	0
	Intr	60403	60569	+	0
	Intr	60679	60921	+	0
	Intr	61199	61287	+	0
20	Intr	61732	61782	+	0
	Intr	61874	61964	+	0
	Intr	62161	62265	+	0
	Intr	62359	62429	+	0
	Intr	62521	62616	+	0
25	Term	62794	63051	+	0

>4519192 /2467

30	len =	407	nex =	1	
	Sngl	23640	24046	+	0

>4519192 /37019

35	len =	2212	nex =	8	
	Init	34427	34741	+	0
	Intr	34950	35008	+	0
	Intr	35106	35180	+	0
40	Intr	35255	35368	+	0
	Intr	35462	35532	+	0
	Intr	35652	35949	+	0
	Intr	36042	36331	+	0
45	Term	36419	36638	+	0

>4519192 /5819

	len =	771	nex =	1	
50	Sngl	3537	2767	-	0

>4519192 /29301

55	len =	2235	nex =	11	
	Init	39652	39840	+	0
	Intr	40077	40138	+	0
	Intr	40278	40380	+	0
	Intr	40499	40557	+	0
60	Intr	40641	40664	+	0

1413

	Intr	40784	40866	+	0
	Intr	40968	41032	+	0
	Intr	41147	41196	+	0
	Intr	41287	41388	+	0
5	Intr	41473	41532	+	0
	Term	41634	41886	+	0

>4519192 /21955

10	len =	2209	nex =	6	
	Term	46248	45906	-	0
	Intr	46484	46326	-	0
	Intr	47033	46716	-	0
15	Intr	47390	47130	-	0
	Intr	47555	47478	-	0
	Init	48114	47633	-	0

>4519192 /533

```

len =      1489    nex =      5
Term 48529 48236 - 0
Intr 48841 48619 - 0
Intr 49138 48966 - 0
Intr 49404 49245 - 0
Init 49705 49484 - 0

```

>4519192 /31447

30	len =	1930	nex =	7	
	Init	76620	76794	+	0
	Intr	77019	77233	+	0
35	Intr	77315	77472	+	0
	Intr	77567	77754	+	0
	Intr	77855	78107	+	0
	Intr	78180	78328	+	0
	Term	78423	78546	+	0

>4519192 /33530

	len =	1896	nex =	6	
45	Init	861	1342	+	0
	Intr	1549	1603	+	0
	Intr	1757	1818	+	0
	Intr	1914	2020	+	0
	Intr	2119	2206	+	0
50	Term	2324	2756	+	0

>4519193 /22588

```

55      len =      1210      nex =          1
      Snq1 10581 11344      +          0

```

>4519193 /12734

```
60      len =      3093      nex =      15
```

1414

	Term	11783	11499	-	0
	Intr	11954	11871	-	0
	Intr	12113	12043	-	0
5	Intr	12282	12213	-	0
	Intr	12459	12367	-	0
	Intr	12649	12556	-	0
	Intr	12798	12737	-	0
10	Intr	12989	12945	-	0
	Intr	13146	13084	-	0
	Intr	13398	13333	-	0
	Intr	13650	13534	-	0
	Intr	13862	13742	-	0
	Intr	14095	13965	-	0
15	Intr	14272	14240	-	0
	Init	14591	14446	-	0
	>4519193	/2934			
20	len =	610	nex =	1	
	Sngl	29932	29326	-	0
	>4519193	/101342			
25	len =	310	nex =	1	
	Sngl	43149	43454	+	0
30	>4519193	/31309			
	len =	763	nex =	1	
	Sngl	51604	50842	-	0
35	>4519193	/38326			
	len =	2324	nex =	6	
40	Term	75991	75658	-	0
	Intr	76453	76077	-	0
	Intr	76876	76829	-	0
	Intr	77135	76973	-	0
	Intr	77307	77229	-	0
45	Init	77981	77783	-	0
	>4519193	/114703			
	len =	576	nex =	1	
50	Sngl	7839	8414	+	0
	>4519194	/20539			
55	len =	1577	nex =	3	
	Init	15272	15482	+	0
	Intr	15570	15615	+	0
	Term	16443	16848	+	0
60					

2025 RELEASE UNDER E.O. 14176

1415

>4519194 /91757

```
len = 1407    nex = 7
```

5	Init	16952	17134	+	0
	Intr	17217	17260	+	0
	Intr	17389	17484	+	0
	Intr	17694	17786	+	0
10	Intr	17919	18006	+	0
	Intr	18095	18156	+	0
	Term	18245	18358	+	0

>4519194 /32257

```
15      len =      910      nex =      1
```

Sngl	41773	40872	-	0
------	-------	-------	---	---

>4519194 /9170

```
len = 1599 nex = 4
```

Term	48741	48223	-	0
------	-------	-------	---	---

Intr	48949	48918	-	0
------	-------	-------	---	---

25	Intr	49210	49075	-	0
----	------	-------	-------	---	---

Init	49821	49647	-	0
------	-------	-------	---	---

>4519194 /36525

```
30      len =      1953      nex =          1
```

Snq1	59919	61346	+	0
------	-------	-------	---	---

>4519194 /7233

```
len = 1270    nex = 4
```

Term	68594	68376	-	0
------	-------	-------	---	---

Intr	68973	68686	-	0
------	-------	-------	---	---

40	Intr	69272	69107	-	0
----	------	-------	-------	---	---

Init	69644	69537	-	0
------	-------	-------	---	---

>4519194 /819

```
45      len =      1270      nex =      4
```

Init	7404	7643	+	0
------	------	------	---	---

Intr	7741	8080	+	0
------	------	------	---	---

Intr	8177	8230	+	0
------	------	------	---	---

Term	8376	8673	+	0
------	------	------	---	---

>4519194 /6181

```
len = 1810    nex = 6
```

Init	74654	74757	+	0
------	-------	-------	---	---

Inter	75098	75146	+	0
-------	-------	-------	---	---

Intr	75472	75579	+	0
------	-------	-------	---	---

Intr	75770	75884	+	0
------	-------	-------	---	---

60	Intr	75966	76083	+	0
----	------	-------	-------	---	---

2000

+ 0

	Term	76205	76459	+	0
	>4519194	/13954			
5	len =	1019	nex =	2	
	Init	9367	9653	+	0
	Term	9736	10385	+	0
10	>4519195	/26815			
	len =	3889	nex =	16	
	Term	12207	11831	-	0
15	Intr	12457	12314	-	0
	Intr	12656	12530	-	0
	Intr	12841	12762	-	0
	Intr	13213	12962	-	0
20	Intr	13440	13312	-	0
	Intr	13706	13657	-	0
	Intr	13944	13915	-	0
	Intr	14169	14056	-	0
	Intr	14435	14374	-	0
25	Intr	14556	14514	-	0
	Intr	14691	14636	-	0
	Intr	14893	14836	-	0
	Intr	15044	14976	-	0
	Intr	15431	15352	-	0
30	Init	15719	15536	-	0
	>4519195	/17362			
	len =	2377	nex =	6	
35	Term	24714	24441	-	0
	Intr	24903	24813	-	0
	Intr	25545	25430	-	0
	Intr	26012	25683	-	0
40	Intr	26589	26090	-	0
	Init	26817	26772	-	0
	>4519195	/40420			
	len =	1753	nex =	6	
45	Init	42376	42569	+	0
	Intr	42664	42895	+	0
	Intr	42979	43142	+	0
	Intr	43253	43535	+	0
50	Intr	43622	43788	+	0
	Term	43875	44128	+	0
	>4519195	/94524			
55	len =	475	nex =	2	
	Init	43654	43788	+	0
	Term	43875	44128	+	0
60	>4519195	/20767			

1417

	len =	1874	nex =	6	
5	Term	47286	46916	-	0
	Intr	47435	47386	-	0
	Intr	47604	47531	-	0
	Intr	48033	47989	-	0
	Intr	48249	48134	-	0
	Init	48789	48477	-	0
10	>4519195 /97350				
	len =	613	nex =	2	
15	Term	67803	67627	-	0
	Init	68239	67907	-	0
	>4519197 /32643				
20	len =	1763	nex =	1	
	Sngl	27667	25905	-	0
	>4521999 /98976				
25	len =	430	nex =	1	
	Sngl	10245	10673	+	0
30	>4521999 /109109				
	len =	399	nex =	1	
	Sngl	10295	10686	+	0
35	>4521999 /106320				
	len =	278	nex =	1	
40	Sngl	10379	10656	+	0
	>4521999 /115957				
	len =	295	nex =	1	
45	Sngl	10445	10739	+	0
	>4521999 /114949				
50	len =	285	nex =	1	
	Sngl	10447	10731	+	0
	>4521999 /10836				
55	len =	562	nex =	1	
	Sngl	12451	11890	-	0
60	>4521999 /151572				

2025 RELEASE UNDER E.O. 14176

1418

	len =	1914	nex =	6	
	Term	22290	22087	-	0
5	Intr	22510	22406	-	0
	Intr	22867	22604	-	0
	Intr	23125	22964	-	0
	Intr	23303	23226	-	0
	Init	24000	23670	-	0
10	>4521999	/31655			
	len =	1630	nex =	5	
15	Term	26057	25794	-	0
	Intr	26219	26143	-	0
	Intr	26450	26332	-	0
	Intr	26899	26820	-	0
	Init	27415	27008	-	0
20	>4521999	/92908			
	len =	2017	nex =	4	
25	Term	31465	31171	-	0
	Intr	31816	31636	-	0
	Intr	32453	32089	-	0
	Init	33187	33057	-	0
30	>4521999	/34062			
	len =	1139	nex =	3	
35	Term	38591	38301	-	0
	Intr	38843	38709	-	0
	Init	39439	39192	-	0
	>4521999	/3948			
40	len =	1096	nex =	1	
	Sngl	9611	9785	+	0
	>4521999	/95453			
45	len =	1044	nex =	3	
	Init	9613	9785	+	0
	Intr	10095	10272	+	0
50	Term	10295	10656	+	0
	>4521999	/17347			
55	len =	1074	nex =	1	
	Sngl	9613	10686	+	0
	>4521999	/125485			
60	len =	1115	nex =	2	

2025 RELEASE UNDER E.O. 14176

1419

	Init	9613	9785	+	0
	Term	10095	10272	+	0
5	>4521999 /125977				
	len =	1120	nex =	3	
10	Init	9613	9785	+	0
	Intr	10095	10272	+	0
	Term	10295	10732	+	0
	>4521999 /98855				
15	len =	1126	nex =	3	
	Init	9613	9785	+	0
	Intr	10095	10272	+	0
	Term	10295	10738	+	0
20	>4521999 /7420				
	len =	986	nex =	2	
25	Init	9615	9785	+	0
	Term	10095	10600	+	0
	>4521999 /234				
30	len =	1118	nex =	2	
	Init	9615	9785	+	0
	Term	10095	10272	+	0
35	>4521999 /26796				
	len =	1119	nex =	3	
40	Init	9615	9785	+	0
	Intr	10095	10272	+	0
	Term	10295	10733	+	0
	>4521999 /13879				
45	len =	1121	nex =	2	
	Init	9615	9785	+	0
	Term	10095	10735	+	0
50	>4522002 /18876				
	len =	2097	nex =	4	
55	Init	31368	31702	+	0
	Intr	31777	31877	+	0
	Intr	32003	32211	+	0
	Term	32776	33464	+	0
60	>4522002 /8550				

1420

	len =	550	nex =	1	
	Sngl	36522	35979	-	0
5	>4522002	/10080			
	len =	2566	nex =	7	
10	Term	37918	37569	-	0
	Intr	38093	38029	-	0
	Intr	38349	38274	-	0
	Intr	38569	38442	-	0
	Intr	38766	38659	-	0
15	Intr	39324	39168	-	0
	Init	40134	39924	-	0
	>4522002	/35371			
20	len =	1570	nex =	3	
	Term	70494	70283	-	0
	Intr	70779	70573	-	0
	Init	71849	71563	-	0
25	>4531433	/142861			
	len =	940	nex =	1	
30	Sngl	33855	32916	-	0
	>4531433	/16421			
	len =	1991	nex =	8	
35	Term	35017	34361	-	0
	Intr	35195	35098	-	0
	Intr	35388	35286	-	0
	Intr	35569	35484	-	0
40	Intr	35771	35684	-	0
	Intr	35976	35863	-	0
	Intr	36161	36073	-	0
	Init	36341	36247	-	0
	>4531433	/41669			
45	len =	276	nex =	1	
	Sngl	37215	36940	-	0
50	>4538895	/7395			
	len =	2470	nex =	6	
55	Term	47375	47093	-	0
	Intr	47611	47463	-	0
	Intr	47817	47700	-	0
	Intr	48199	48007	-	0
	Intr	48407	48286	-	0
60	Init	49561	49113	-	0

2025 RELEASE UNDER E.O. 14176

1421

	>4538895	/8795		
	len =	756	nex =	1
5	Sngl	49466	48947	- 0
	>4538895	/24071		
	len =	1392	nex =	3
10	Init	72494	72654	+ 0
	Intr	72892	73080	+ 0
	Term	73174	73422	+ 0
15	>4538895	/27800		
	len =	790	nex =	3
	Init	72493	72654	+ 0
20	Intr	72892	73080	+ 0
	Term	73174	73280	+ 0
	>4538895	/19582		
25	len =	1899	nex =	2
	Term	6321	5522	- 0
	Init	7420	6901	- 0
30	>4538895	/25257		
	len =	1939	nex =	2
	Term	6321	5516	- 0
35	Init	7454	6901	- 0
	>4538895	/142381		
	len =	1954	nex =	3
40	Init	87087	87340	+ 0
	Intr	88405	88485	+ 0
	Term	88572	89040	+ 0
45	>4538918	/33495		
	len =	2564	nex =	6
	Init	35043	35417	+ 0
50	Intr	35652	35904	+ 0
	Intr	36064	36212	+ 0
	Intr	36295	36552	+ 0
	Intr	36711	36902	+ 0
	Term	37042	37606	+ 0
55	>4538918	/15410		
	len =	1705	nex =	4
60	Term	52803	52534	- 0

2025 RELEASE UNDER E.O. 14176

1422

Intr	53006	52889	-	0
Intr	53214	53100	-	0
Init	53440	53336	-	0

5 >4538918 /31259

len = 2096 nex = 7

10	Init	55471	55645	+	0
	Intr	56096	56230	+	0
	Intr	56317	56541	+	0
	Intr	56641	56823	+	0
	Intr	56932	57155	+	0
	Intr	57239	57284	+	0
15	Term	57378	57566	+	0

>4538918 /34934

20 len = 3130 nex = 12

	Init	70373	70535	+	0
	Intr	70658	70701	+	0
	Intr	70897	70992	+	0
	Intr	71153	71245	+	0
25	Intr	71398	71488	+	0
	Intr	71582	71643	+	0
	Intr	71733	71917	+	0
	Intr	72056	72233	+	0
	Intr	72312	72355	+	0
30	Intr	72493	72682	+	0
	Intr	72788	72805	+	0
	Term	73209	73502	+	0

>4538918 /7421

35 len = 1241 nex = 1

Sngl 79061 77821 - 0

40 >4538918 /148597

len = 1482 nex = 4

45	Init	93052	93292	+	0
	Intr	93665	93847	+	0
	Intr	94012	94176	+	0
	Term	94312	94533	+	0

>4538949 /13231

50 len = 738 nex = 3

	Init	52100	52260	+	0
	Intr	52368	52518	+	0
55	Term	52605	52837	+	0

>4538949 /39765

60 len = 1486 nex = 6

1423

	Init	53668	53951	+	0
	Intr	54035	54119	+	0
	Intr	54209	54259	+	0
	Intr	54358	54478	+	0
5	Intr	54585	54719	+	0
	Term	54798	55153	+	0
	>4538949 /93802				
10	len =	310	nex =	1	
	Sngl	54895	55195	+	0
	>4538949 /1354				
15	len =	1825	nex =	7	
	Init	66408	66496	+	0
	Intr	66590	66728	+	0
20	Intr	66808	66882	+	0
	Intr	67001	67122	+	0
	Intr	67205	67280	+	0
	Intr	67362	67421	+	0
	Term	67521	67843	+	0
25	>4538949 /36655				
	len =	3584	nex =	4	
30	Term	77287	76801	-	0
	Intr	77426	77365	-	0
	Intr	78110	77691	-	0
	Init	80384	79836	-	0
35	>4538972 /6822				
	len =	2801	nex =	13	
	Term	11535	11345	-	0
40	Intr	11698	11621	-	0
	Intr	11841	11769	-	0
	Intr	12019	11939	-	0
	Intr	12214	12164	-	0
	Intr	12391	12297	-	0
45	Intr	12711	12631	-	0
	Intr	12872	12788	-	0
	Intr	13313	13234	-	0
	Intr	13530	13416	-	0
	Intr	13703	13633	-	0
50	Intr	13870	13786	-	0
	Init	14145	13945	-	0
	>4538972 /41992				
55	len =	2052	nex =	5	
	Init	32316	32561	+	0
	Intr	32643	32715	+	0
	Intr	33489	33697	+	0
60	Intr	33789	33935	+	0

2025-09-25 14:23:00

	Term	34034	34367	+	1424	0
	>4538972	/35962				
5	len =	823	nex =	1		
	Sngl	35958	36389	+		0
	>4538972	/31115				
10	len =	1478	nex =	3		
	Init	3879	4068	+		0
	Intr	4258	4391	+		0
15	Term	5063	5356	+		0
	>4538990	/32791				
	len =	4390	nex =	8		
20	Term	22701	22300	-		0
	Intr	22820	22779	-		0
	Intr	22937	22896	-		0
	Intr	23128	23029	-		0
25	Intr	23284	23223	-		0
	Intr	23456	23375	-		0
	Intr	25216	25034	-		0
	Init	26686	26545	-		0
30	>4538990	/36491				
	len =	533	nex =	1		
	Sngl	4200	4732	+		0
35	>4538990	/40182				
	len =	2574	nex =	8		
40	Term	47130	47007	-		0
	Intr	47313	47244	-		0
	Intr	47449	47400	-		0
	Intr	47627	47560	-		0
	Intr	47778	47717	-		0
45	Intr	47991	47868	-		0
	Intr	48467	48429	-		0
	Init	48934	48614	-		0
	>4538990	/6718				
50	len =	2984	nex =	12		
	Term	56364	56049	-		0
	Intr	56633	56462	-		0
55	Intr	56834	56722	-		0
	Intr	57028	56963	-		0
	Intr	57314	57213	-		0
	Intr	57443	57391	-		0
	Intr	57592	57535	-		0
60	Intr	57780	57682	-		0

2025-06-06 09:40:40

1425

	Intr	57898	57855	-	0
	Intr	58299	58209	-	0
	Intr	58481	58455	-	0
	Init	59032	58791	-	0
5	>4539290 /9546				
	len =	1353	nex =	3	
10	Init	31019	31261	+	0
	Intr	31721	31841	+	0
	Term	31919	32371	+	0
	>4539290 /111551				
15	len =	680	nex =	2	
	Init	31692	31841	+	0
	Term	31919	32371	+	0
20	>4539290 /98650				
	len =	1931	nex =	5	
25	Init	36398	36658	+	0
	Intr	36936	37160	+	0
	Intr	37542	37689	+	0
	Intr	37807	37886	+	0
	Term	37970	38328	+	0
30	>4539290 /6966				
	len =	955	nex =	2	
35	Term	57171	56874	-	0
	Init	57828	57579	-	0
	>4539290 /16673				
40	len =	976	nex =	1	
	Sngl	57850	56875	-	0
	>4539290 /31982				
45	len =	1005	nex =	2	
	Term	57171	56874	-	0
	Init	57878	57579	-	0
50	>4539290 /37817				
	len =	2214	nex =	8	
55	Init	58762	59031	+	0
	Intr	59151	59225	+	0
	Intr	59572	59628	+	0
	Intr	59712	59866	+	0
	Intr	59985	60119	+	0
60	Intr	60211	60313	+	0

1426

	Intr	60402	60501	+	0
	Term	60611	60975	+	0
	>4539290 /103513				
5	len =	198	nex =	1	
	Sngl	60820	61017	+	0
10	>4539290 /10511				
	len =	1487	nex =	5	
	Term	4951	4724	-	0
15	Intr	5226	5106	-	0
	Intr	5588	5309	-	0
	Intr	6021	5973	-	0
	Init	6210	6132	-	0
20	>4539290 /24045				
	len =	713	nex =	2	
	Init	86190	86240	+	0
25	Term	86397	86784	+	0
	>4539309 /31833				
	len =	1938	nex =	3	
30	Init	23367	23550	+	0
	Intr	24243	24528	+	0
	Term	24866	25304	+	0
35	>4539309 /40069				
	len =	1415	nex =	1	
	Sngl	28842	30256	+	0
40	>4539309 /119970				
	len =	1487	nex =	6	
45	Term	40137	39707	-	0
	Intr	40406	40272	-	0
	Intr	40622	40484	-	0
	Intr	40793	40703	-	0
	Intr	41003	40894	-	0
50	Init	41193	41096	-	0
	>4539309 /6351				
	len =	2193	nex =	7	
55	Term	40137	39693	-	0
	Intr	40406	40272	-	0
	Intr	40622	40484	-	0
	Intr	40793	40703	-	0
60	Intr	41003	40894	-	0

1427

	Intr	41365	41096	-	0
	Init	41885	41628	-	0
5	>4539309	/42753			
	len =	911	nex =	3	
	Term	65462	65294	-	0
10	Intr	65761	65543	-	0
	Init	66204	66078	-	0
	>4539309	/18408			
15	len =	1111	nex =	2	
	Term	65761	65543	-	0
	Init	66377	66078	-	0
20	>4539309	/28021			
	len =	1427	nex =	2	
	Term	65761	65543	-	0
25	Init	66693	66078	-	0
	>4539309	/6045			
	len =	262	nex =	1	
30	Sngl	92055	91794	-	0
	>4539309	/8278			
35	len =	2297	nex =	5	
	Term	94508	94131	-	0
	Intr	94898	94766	-	0
	Intr	95549	95481	-	0
40	Intr	96124	95779	-	0
	Init	96427	96207	-	0
	>4539331	/38915			
45	len =	790	nex =	3	
	Term	47037	46932	-	0
	Intr	47423	47129	-	0
	Init	47720	47526	-	0
50	>4539353	/7769			
	len =	556	nex =	1	
	Sngl	2681	2129	-	0
55	>4539353	/92155			
	len =	411	nex =	1	
60	Sngl	2685	2275	-	0

1428

>4539353 /1732

5	len =	503	nex =	1	
	Sngl	53207	53709	+	0

>4539378 /2917

10	len =	370	nex =	1	
	Sngl	34714	34349	-	0

>4539378 /2359

15	len =	1179	nex =	5	
	Init	61039	61098	+	0
	Intr	61178	61285	+	0
20	Intr	61443	61485	+	0
	Intr	61581	61776	+	0
	Term	61866	62217	+	0

>4539378 /38408

25	len =	1035	nex =	2	
	Term	72291	71647	-	0
	Init	72681	72375	-	0

>4539378 /7322

30	len =	1615	nex =	2	
35	Term	79437	78787	-	0
	Init	80401	79977	-	0

>4539402 /8976

40	len =	2875	nex =	10	
	Term	27509	27384	-	0
	Intr	27711	27643	-	0
45	Intr	28151	28065	-	0
	Intr	28346	28239	-	0
	Intr	28516	28426	-	0
	Intr	29461	29382	-	0
	Intr	29655	29551	-	0
50	Intr	29831	29737	-	0
	Intr	29988	29910	-	0
	Init	30258	30076	-	0

>4539402 /41359

55	len =	3130	nex =	10	
	Term	27509	27214	-	0
	Intr	27711	27643	-	0
	Intr	28151	28065	-	0
60	Intr	28346	28239	-	0

2025-09-05 16:00:00

1429

	Intr	28516	28426	-	0
	Intr	29461	29382	-	0
	Intr	29655	29551	-	0
	Intr	29831	29737	-	0
5	Intr	29988	29910	-	0
	Init	30343	30076	-	0
	>4539402		/41155		
10	len =	918	nex =	2	
	Term	35555	35468	-	0
	Init	36257	36135	-	0
15	>4539402		/21228		
	len =	1469	nex =	3	
	Term	35555	35325	-	0
20	Intr	36257	36135	-	0
	Init	36793	36352	-	0
	>4539415		/742		
25	len =	1403	nex =	3	
	Init	26047	26138	+	0
	Intr	26513	26661	+	0
	Term	27042	27449	+	0
30	>4539415		/14013		
	len =	828	nex =	1	
35	Sngl	30384	30766	+	0
	>4539415		/14312		
40	len =	2186	nex =	6	
	Term	31374	31324	-	0
	Intr	31712	31634	-	0
	Intr	32178	32124	-	0
	Intr	32659	32576	-	0
45	Intr	32806	32745	-	0
	Init	33072	32898	-	0
	>4539448		/38976		
50	len =	401	nex =	1	
	Sngl	19914	19514	-	0
	>4539448		/13796		
55	len =	1156	nex =	2	
	Term	19237	18760	-	0
	Init	19915	19490	-	0
60					

005555-00460

1430

>4539448

/8846

len = 1419 nex = 6

5	Term	3827	3621	-	0
	Intr	4092	3925	-	0
	Intr	4400	4282	-	0
	Intr	4556	4495	-	0
	Intr	4723	4647	-	0
10	Init	5039	4865	-	0

>4539448

/26831

len = 1511 nex = 6

15	Term	3827	3616	-	0
	Intr	4092	3925	-	0
	Intr	4400	4282	-	0
	Intr	4556	4495	-	0
20	Intr	4723	4647	-	0
	Init	5126	4865	-	0

>4544365

/32729

len = 1218 nex = 1

Sngl 13401 14618 + 0

>4544365

/3839

len = 2110 nex = 3

30	Term	23021	22565	-	0
	Intr	23867	23630	-	0
35	Init	24665	24122	-	0

>4544365

/32380

len = 1118 nex = 2

40	Term	44571	44182	-	0
	Init	45299	44761	-	0

>4544365

/28318

len = 970 nex = 2

	Init	47075	47349	+	0
	Term	47442	48044	+	0

50					
----	--	--	--	--	--

>4544365

/9657

len = 533 nex = 1

55	Sngl	80944	80412	-	0
----	------	-------	-------	---	---

>4544381

/30607

len = 1011 nex = 3

60					
----	--	--	--	--	--

2025 RELEASE UNDER E.O. 14176

1431

Init	43399	43605	+	0
Intr	43651	43747	+	0
Term	43844	44402	+	0

5 >4544381 /38904

len =	207	nex =	1	
-------	-----	-------	---	--

10 Sngl 44129 44335 + 0

>4544381 /6922

len =	827	nex =	1	
-------	-----	-------	---	--

15 Sngl 93744 92918 - 0

>4544381 /43076

len =	1855	nex =	1	
-------	------	-------	---	--

Sngl	94771	92917	-	0
------	-------	-------	---	---

>4544405 /3366

25 len = 1270 nex = 6

Init	108883	109010	+	0
------	--------	--------	---	---

Intr	109298	109331	+	0
------	--------	--------	---	---

Intr	109412	109584	+	0
------	--------	--------	---	---

30 Intr 109671 109722 + 0

Intr	109823	109890	+	0
------	--------	--------	---	---

Term	109982	110152	+	0
------	--------	--------	---	---

>4544405 /21909

35 len = 1781 nex = 6

Init	25216	25418	+	0
------	-------	-------	---	---

Intr	25776	25847	+	0
------	-------	-------	---	---

40 Intr 26206 26284 + 0

Intr	26395	26527	+	0
------	-------	-------	---	---

Intr	26623	26661	+	0
------	-------	-------	---	---

Term	26747	26996	+	0
------	-------	-------	---	---

45 >4544405 /11114

len =	1439	nex =	1	
-------	------	-------	---	--

Sngl	51750	50803	-	0
------	-------	-------	---	---

50 >4544405 /40781

len =	1633	nex =	6	
-------	------	-------	---	--

55 Term 80918 80716 - 0

Intr	81127	81067	-	0
------	-------	-------	---	---

Intr	81666	81595	-	0
------	-------	-------	---	---

Intr	81789	81740	-	0
------	-------	-------	---	---

Intr	82070	81958	-	0
------	-------	-------	---	---

60 Init 82348 82159 - 0

1432

>4544405 /104918

```

5      len =      1161    nex =          1
      Sngl  90572    90694      +          0

```

>4544435 /20723

```
10      len =      1413      nex =          1
        Sngl 29535  30947      +          0
```

>4544435 /111553

```
len =      510    nex =      1
Sngl  30533    31042      +      0
```

20 >4544435 /117955

```
len =      1450      nex =          3

Init      3407      3589      +      0
Intr      4304      4489      +      0
Term      4602      4848      +      0
```

>4544435 /32573

30	len =	3571	nex =	5	
	Term	49696	49307	-	0
	Intr	50351	50080	-	0
	Intr	50691	50599	-	0
35	Intr	52326	52225	-	0
	Init	52877	52427	-	

>4544435 /38277

```

40      len =      535      nex =          1
      Sngl  78881  78813      -          0

```

45 >4544435 /26961

len =	1350	nex =	5	
Init	92102	92249	+	0
Intr	92397	92496	+	0
Intr	92587	92644	+	0
Intr	92722	93078	+	0
Term	93165	93451	+	0

>4544435 /10312

```
len =      2925      nex =      6
Term  94666  93767    -    0
Intr  94855  94772    -    0
Intr  94981  94933    -    0
```

[illegible]

1433

Intr	95346	95304	-	0
Intr	95580	95523	-	0
Init	96143	96015	-	0

5 >4557056 /38982

len = 1316 nex = 4

Term	13849	13681	-	0
Intr	14205	14054	-	0
Intr	14347	14277	-	0
Init	14637	14454	-	0

>4557061 /7520

15 len = 1480 nex = 1

Sngl 23980 24135 + 0

20 >4557061 /39407

len = 1630 nex = 3

Term	26056	25532	-	0
Intr	26195	26155	-	0
Init	27156	26894	-	0

>4557061 /5435

30 len = 1632 nex = 3

Term	26056	25532	-	0
Intr	26195	26155	-	0
Init	27012	26894	-	0

35 >4558521 /23664

len = 898 nex = 1

40 Sngl 30777 31674 + 0

>4558521 /37480

len = 760 nex = 2

Term	41994	41798	-	0
Init	42557	42246	-	0

>4558521 /27813

50 len = 1810 nex = 4

Term	60751	60244	-	0
Intr	61451	61243	-	0
Intr	61676	61576	-	0
Init	62053	61830	-	0

>4558521 /7488

60 len = 1414 nex = 2

000100-02000000

1434

Term	83600	83130	-	0
Init	84090	83849	-	0

5 >4558586 /18981

```
len =      3812      nex =      12
```

	Term	7302	6978	-	0
10	Intr	7549	7499	-	0
	Intr	7962	7870	-	0
	Intr	8232	8157	-	0
	Intr	8456	8340	-	0
15	Intr	8680	8616	-	0
	Intr	9112	8999	-	0
	Intr	9310	9212	-	0
	Intr	9563	9472	-	0
20	Intr	10014	9891	-	0
	Intr	10285	10118	-	0
	Init	10789	10619	-	0

>4558586 /18724

25	len =	4210	nex =	14	
	Term	11491	11046	-	0
	Intr	11945	11761	-	0
	Intr	12171	12058	-	0
30	Intr	12446	12359	-	0
	Intr	12612	12540	-	0
	Intr	13246	13121	-	0
	Intr	13516	13427	-	0
	Intr	13708	13598	-	0
	Intr	14034	13801	-	0
35	Intr	14205	14128	-	0
	Intr	14379	14308	-	0
	Intr	14551	14468	-	0
	Intr	14778	14659	-	0
	Init	15254	15074	-	0

>4558586 /2070

```
len =      1824      nex =          7
```

45	Term	33890	33632	-	0
	Intr	34060	33981	-	0
	Intr	34417	34262	-	0
	Intr	34645	34510	-	0
	Intr	34881	34730	-	0
50	Intr	35164	35075	-	0
	Init	35455	35254	-	0

>4558586 /157873

```

55      len =      649      nex =      3
      Term 34881  34807      -      0
      Intr 35164  35075      -      0
      Init 35455  35254      -      0

```

1435

>4558586 /35814

len = 1938 nex = 7

5	Term	33890	33575	-	0
	Intr	34060	33981	-	0
	Intr	34417	34262	-	0
	Intr	34645	34510	-	0
	Intr	34881	34730	-	0
10	Intr	35164	35075	-	0
	Init	35512	35254	-	0

>4558586 /19277

15 len = 655 nex = 1

Sngl 77232 77886 + 0

>4558586 /106135

20 len = 635 nex = 1

Sngl 84401 85035 + 0

25 >4558590 /41858

len = 2113 nex = 3

30	Init	3306	3445	+	0
	Intr	3564	4886	+	0
	Term	5194	5418	+	0

>4558590 /14128

35 len = 1425 nex = 2

	Init	3948	4886	+	0
	Term	5194	5372	+	0

40 >4558656 /95747

len = 859 nex = 1

45 Sngl 82575 83433 + 0

>4558656 /25241

len = 657 nex = 1

50 Sngl 82776 83432 + 0

>4558656 /17483

len = 2449 nex = 2

55	Term	98045	96936	-	0
	Init	99384	98603	-	0

>4558674 /110217

60

1437

>4559319 /37057

5	len =	3250	nex =	12	
	Init	84481	84692	+	0
	Intr	85229	85264	+	0
	Intr	85302	85358	+	0
	Intr	85490	85573	+	0
10	Intr	85652	85723	+	0
	Intr	85920	85991	+	0
	Intr	86232	86272	+	0
	Intr	86584	86650	+	0
	Intr	86992	87183	+	0
15	Intr	87283	87372	+	0
	Intr	87461	87568	+	0
	Term	87662	87727	+	0

>4559344 /17434

20	len =	1227	nex =	2	
	Term	99351	98868	-	0
	Init	100094	99455	-	0

>4559344 /30071

	len =	1657	nex =	5	
30	Term	36166	35586	-	0
	Intr	36370	36254	-	0
	Intr	36762	36454	-	0
	Intr	37029	36967	-	0
35	Init	37242	37124	-	0

>4559344 /10798

	len =	1189	nex =	5	
40	Init	45685	45867	+	0
	Intr	45942	46037	+	0
	Intr	46237	46359	+	0
	Intr	46569	46608	+	0
45	Term	46704	46873	+	0

>4559344 /1916

	len =	490	nex =	1	
50	Sngl	57577	57090	-	0

>4559344 /6722

	len =	409	nex =	1	
55	Sngl	57584	57176	-	0

>4559344 /10261

60	len =	522	nex =	1	
----	-------	-----	-------	---	--

1438

	Sngl	57598	57077	-	0
	>4559344	/97914			
5	len =	1291	nex =	3	
	Init	78733	78958	+	0
	Intr	79545	79694	+	0
10	Term	79797	80023	+	0
	>4559344	/14570			
	len =	2280	nex =	7	
15	Init	81776	82132	+	0
	Intr	82307	82649	+	0
	Intr	82774	82992	+	0
	Intr	83092	83151	+	0
20	Intr	83231	83386	+	0
	Intr	83500	83593	+	0
	Term	83750	84055	+	0
	>4559344	/122569			
25	len =	214	nex =	1	
	Sngl	83842	84055	+	0
30	>4559344	/36270			
	len =	1377	nex =	1	
	Sngl	91284	90056	-	0
35	>4559344	/29774			
	len =	1937	nex =	3	
40	Term	91284	90056	-	0
	Intr	91468	91366	-	0
	Init	91992	91560	-	0
	>4559375	/102981			
45	len =	568	nex =	1	
	Sngl	19716	20275	+	0
50	>4559375	/94104			
	len =	516	nex =	1	
	Sngl	21573	22085	+	0
55	>4559375	/12474			
	len =	737	nex =	1	
60	Sngl	26746	26629	-	0

2025 RELEASE UNDER E.O. 14176

1439

>4559375 /35220

5	len =	1760	nex =	7	
	Init	38581	38611	+	0
	Intr	38834	39022	+	0
	Intr	39074	39179	+	0
	Intr	39265	39450	+	0
10	Intr	39547	39712	+	0
	Intr	39807	40123	+	0
	Term	40251	40340	+	0

>4567193 /38340

15	len =	1426	nex =	4	
	Term	16701	16432	-	0
	Intr	17094	16777	-	0
20	Intr	17731	17187	-	0
	Init	17857	17801	-	0

>4567193 /17172

25	len =	1760	nex =	6	
	Term	28035	27578	-	0
	Intr	28415	28113	-	0
	Intr	28602	28501	-	0
30	Intr	28796	28685	-	0
	Intr	29087	29011	-	0
	Init	29337	29173	-	0

>4567193 /206273

35	len =	2060	nex =	7	
	Term	30185	29858	-	0
	Intr	30460	30268	-	0
40	Intr	30657	30541	-	0
	Intr	31111	30896	-	0
	Intr	31336	31225	-	0
	Intr	31570	31506	-	0
45	Init	31737	31656	-	0

>4567193 /15405

	len =	192	nex =	1	
50	Sngl	39059	39250	+	0

>4567193 /30185

	len =	1632	nex =	4	
55	Term	41869	41427	-	0
	Intr	42547	42403	-	0
	Intr	42721	42661	-	0
60	Init	43058	42799	-	0

20250101 09:22:56

1440

>4567193 /15081

len = 928 nex = 2

5	Init	65231	65374	+	0
	Term	65556	66158	+	0

>4567193 /153017

10 len = 616 nex = 2

	Init	65262	65374	+	0
	Term	65556	65877	+	0

15 >4567237 /13309

len = 1334 nex = 5

20	Init	37681	37817	+	0
	Intr	38008	38054	+	0
	Intr	38602	38657	+	0
	Intr	38750	38866	+	0
	Term	38969	39006	+	0

25 >4567237 /25127

len = 2314 nex = 8

30	Init	52101	52260	+	0
	Intr	52608	52692	+	0
	Intr	52766	52834	+	0
	Intr	52909	53021	+	0
	Intr	53119	53230	+	0
35	Intr	53328	53471	+	0
	Intr	53682	53856	+	0
	Term	54076	54414	+	0

>4567237 /22703

40 len = 1450 nex = 3

	Term	56329	56148	-	0
	Intr	56617	56415	-	0
	Init	57116	56959	-	0

45 >4567237 /33370

len = 685 nex = 3

50	Term	76509	76290	-	0
	Intr	76701	76598	-	0
	Init	76968	76910	-	0

>4567237 /11284

55 len = 693 nex = 3

	Term	76509	76285	-	0
	Intr	76701	76598	-	0

60 Init 76968 76910 - 0

2025 RELEASE UNDER E.O. 14176

1441

>4567259 /33

5	len =	1944	nex =	6	
	Term	678	284	-	0
	Intr	980	761	-	0
	Intr	1195	1063	-	0
	Intr	1445	1328	-	0
10	Intr	1621	1527	-	0
	Init	2227	2024	-	0

>4567259 /38112

15	len =	1572	nex =	5	
	Term	26086	25917	-	0
	Intr	26263	26173	-	0
	Intr	26465	26356	-	0
20	Intr	26929	26660	-	0
	Init	27488	27013	-	0

>4567259 /20834

25	len =	3815	nex =	14	
	Init	29608	30133	+	0
	Intr	30762	30856	+	0
30	Intr	30987	31095	+	0
	Intr	31188	31234	+	0
	Intr	31412	31467	+	0
	Intr	31629	31767	+	0
	Intr	31884	32021	+	0
	Intr	32100	32168	+	0
35	Intr	32254	32364	+	0
	Intr	32467	32595	+	0
	Intr	32688	32765	+	0
	Intr	32854	32925	+	0
	Intr	33026	33088	+	0
40	Term	33177	33422	+	0

>4567259 /121118

45	len =	2034	nex =	9	
	Init	31629	31767	+	0
	Intr	31884	32021	+	0
	Intr	32100	32168	+	0
50	Intr	32254	32364	+	0
	Intr	32467	32595	+	0
	Intr	32688	32765	+	0
	Intr	32854	32925	+	0
	Intr	33026	33088	+	0
	Term	33177	33474	+	0

>4567259 /21730

```

len =      1774    nex =          6
60    Term      2963    2647      -      0

```

1442

	Intr	3263	3044	-	0
	Intr	3478	3346	-	0
	Intr	3746	3629	-	0
	Intr	3925	3831	-	0
5	Init	4420	4216	-	0
	>4567259 /34470				
	len =	2606	nex =	8	
10	Init	5119	5168	+	0
	Intr	5278	5355	+	0
	Intr	5522	5546	+	0
	Intr	5632	5664	+	0
15	Intr	5971	6101	+	0
	Intr	6291	6511	+	0
	Intr	6606	6853	+	0
	Term	6954	7195	+	0
20	>4567259 /33121				
	len =	2209	nex =	9	
	Init	46682	46836	+	0
25	Intr	47131	47236	+	0
	Intr	47341	47382	+	0
	Intr	47532	47597	+	0
	Intr	47686	47802	+	0
	Intr	47939	48214	+	0
30	Intr	48303	48350	+	0
	Intr	48427	48548	+	0
	Term	48699	48890	+	0
	>4567259 /36598				
35	len =	2212	nex =	11	
	Init	46682	46836	+	0
	Intr	46924	47015	+	0
40	Intr	47131	47236	+	0
	Intr	47341	47382	+	0
	Intr	47532	47597	+	0
	Intr	47686	47802	+	0
	Intr	47939	48013	+	0
45	Intr	48112	48214	+	0
	Intr	48303	48350	+	0
	Intr	48427	48548	+	0
	Term	48699	48893	+	0
50	>4567259 /109948				
	len =	475	nex =	2	
	Init	48460	48548	+	0
55	Term	48699	48934	+	0
	>4567259 /1285				
60	len =	1185	nex =	1	

1443

	Sngl	52746	53930	+	0
	>4567259	/21453			
5	len =	1410	nex =	3	
	Term	69518	68949	-	0
	Intr	69807	69688	-	0
	Init	70358	69891	-	0
10	>4567259	/33656			
	len =	463	nex =	1	
15	Sngl	9884	10346	+	0
	>4567259	/156111			
	len =	322	nex =	1	
20	Sngl	9910	10231	+	0
	>4567300	/125386			
25	len =	910	nex =	3	
	Init	13167	13335	+	0
	Intr	13426	13542	+	0
	Term	13632	14051	+	0
30	>4567300	/101253			
	len =	859	nex =	3	
35	Init	33481	33676	+	0
	Intr	33813	33938	+	0
	Term	34031	34339	+	0
	>4567300	/123234			
40	len =	1330	nex =	4	
	Term	34776	34516	-	0
	Intr	35061	34867	-	0
45	Intr	35283	35188	-	0
	Init	35576	35463	-	0
	>4567300	/24221			
50	len =	393	nex =	1	
	Sngl	8449	8841	+	0
	>4572664	/2867			
55	len =	1510	nex =	4	
	Term	55509	55260	-	0
	Intr	55693	55605	-	0
60	Intr	55882	55783	-	0

2025 RELEASE UNDER E.O. 14176

1444

	Init	56077	56005	-	0
	>4572664	/34807			
5	len =	1418	nex =	1	
	Sngl	59430	60847	+	0
	>4572664	/113985			
10	len =	173	nex =	1	
	Sngl	66412	66584	+	0
	>4572664	/11953			
	len =	1510	nex =	3	
	Init	66481	67144	+	0
20	Intr	67636	67764	+	0
	Term	67855	67987	+	0
	>4572664	/34681			
25	len =	1379	nex =	2	
	Term	73338	72795	-	0
	Init	74173	73545	-	0
	>4572664	/37401			
30	len =	1316	nex =	2	
	Init	76483	76979	+	0
35	Term	77234	77798	+	0
	>4572664	/36719			
	len =	1231	nex =	1	
40	Sngl	79365	78135	-	0
	>4580365	/27596			
45	len =	1365	nex =	4	
	Term	99273	99000	-	0
	Intr	99825	99644	-	0
	Intr	100012	99954	-	0
50	Init	100364	100196	-	0
	>4580365	/27081			
	len =	492	nex =	1	
55	Sngl	43711	43220	-	0
	>4580365	/39571			
60	len =	2540	nex =	9	

000000-000000

1445

	Init	60690	60861	+	0
	Intr	61125	61240	+	0
	Intr	61348	61407	+	0
5	Intr	61645	61830	+	0
	Intr	61925	62029	+	0
	Intr	62147	62243	+	0
	Intr	62438	62545	+	0
	Intr	62686	62821	+	0
10	Term	62943	63229	+	0
>4580365		/108216			
15	len =	1090	nex =	4	
	Init	62147	62243	+	0
	Intr	62438	62545	+	0
	Intr	62686	62821	+	0
	Term	62943	63229	+	0
20	>4580365	/158765			
	len =	2230	nex =	4	
	Init	96491	97832	+	0
25	Intr	97943	98008	+	0
	Intr	98083	98148	+	0
	Term	98215	98259	+	0
>4580365		/3900			
30	len =	1249	nex =	5	
	Init	97546	97832	+	0
	Intr	97943	98008	+	0
35	Intr	98083	98148	+	0
	Intr	98215	98259	+	0
	Term	98386	98794	+	0
>4580454		/36815			
40	len =	2248	nex =	6	
	Term	23978	23775	-	0
	Intr	24212	24072	-	0
45	Intr	24520	24304	-	0
	Intr	24794	24625	-	0
	Intr	25442	25197	-	0
	Init	26022	25523	-	0
>4580454		/12246			
50	len =	1961	nex =	7	
	Init	26312	26626	+	0
55	Intr	26975	27027	+	0
	Intr	27104	27213	+	0
	Intr	27296	27463	+	0
	Intr	27553	27650	+	0
	Intr	27792	27918	+	0
60	Term	28033	28272	+	0

2025 RELEASE UNDER E.O. 14176

1446

>4580454 /123469

5	len =	2112	nex =	7	
	Init	29262	29569	+	0
	Intr	30006	30058	+	0
	Intr	30153	30262	+	0
	Intr	30351	30518	+	0
10	Intr	30608	30705	+	0
	Intr	30832	30958	+	0
	Term	31073	31373	+	0

>4580732 /43073

15	len =	1642	nex =	1	
	Sngl	102136	103777	+	0

20 >4580732 /1069

	len =	1112	nex =	3	
	Term	42616	42417	-	0
25	Intr	43115	43025	-	0
	Init	43528	43375	-	0

>4580744 /34680

30	len =	2025	nex =	4	
	Term	15210	14726	-	0
	Intr	15766	15298	-	0
	Intr	16116	15998	-	0
35	Init	16750	16441	-	0

>4580744 /1294

40	len =	2050	nex =	5	
	Term	22434	22008	-	0
	Intr	22792	22540	-	0
	Intr	23073	22873	-	0
	Intr	23265	23147	-	0
45	Init	24057	23791	-	0

>4580744 /6124

50	len =	251	nex =	1	
	Sngl	31321	31071	-	0

>4580744 /36505

55	len =	3231	nex =	1	
	Sngl	32321	31071	-	0

60	>4580744	/97415			
----	----------	--------	--	--	--

1447

	len =	854	nex =	1	
	Sngl	40155	39760	-	0
5	>4580744	/117519			
	len =	1292	nex =	1	
10	Sngl	41250	42541	+	0
	>4580744	/783			
	len =	1575	nex =	5	
15	Term	44523	44000	-	0
	Intr	44831	44655	-	0
	Intr	45013	44928	-	0
	Intr	45231	45112	-	0
	Init	45559	45529	-	0
20	>4580744	/37712			
	len =	2875	nex =	12	
25	Term	44523	44251	-	0
	Intr	44831	44655	-	0
	Intr	45013	44928	-	0
	Intr	45231	45112	-	0
	Intr	45559	45529	-	0
30	Intr	45722	45647	-	0
	Intr	45995	45908	-	0
	Intr	46153	46092	-	0
	Intr	46314	46238	-	0
	Intr	46521	46419	-	0
35	Intr	46677	46616	-	0
	Init	47125	47027	-	0
	>4580744	/37965			
40	len =	3581	nex =	14	
	Term	55823	55611	-	0
	Intr	56063	55914	-	0
	Intr	56251	56154	-	0
45	Intr	56471	56331	-	0
	Intr	56734	56585	-	0
	Intr	56994	56818	-	0
	Intr	57262	57082	-	0
	Intr	57604	57559	-	0
50	Intr	57984	57921	-	0
	Intr	58092	58070	-	0
	Intr	58311	58239	-	0
	Intr	58627	58496	-	0
	Intr	58799	58713	-	0
55	Init	59191	59053	-	0
	>4580744	/40575			
60	len =	3683	nex =	14	

1448

	Term	55823	55622	-	0
	Intr	56063	55914	-	0
	Intr	56251	56154	-	0
5	Intr	56471	56331	-	0
	Intr	56734	56585	-	0
	Intr	56994	56818	-	0
	Intr	57262	57082	-	0
	Intr	57604	57559	-	0
10	Intr	57984	57921	-	0
	Intr	58092	58070	-	0
	Intr	58311	58239	-	0
	Intr	58627	58496	-	0
	Intr	58799	58713	-	0
	Init	59304	59053	-	0
15	>4580744	/28624			
	len =	2278	nex =	6	
20	Init	60438	60505	+	0
	Intr	60760	60832	+	0
	Intr	61116	61231	+	0
	Intr	61311	61403	+	0
	Intr	61511	61610	+	0
25	Term	61781	62029	+	0
	>4580744	/94743			
	len =	463	nex =	2	
30	Init	67354	67484	+	0
	Term	67496	67816	+	0
	>4580744	/39230			
35	len =	920	nex =	2	
	Term	74699	74303	-	0
	Init	75222	74942	-	0
40	>4580744	/18908			
	len =	2052	nex =	6	
45	Init	75765	76184	+	0
	Intr	76412	76510	+	0
	Intr	76599	76760	+	0
	Intr	76852	76973	+	0
	Intr	77156	77275	+	0
50	Term	77360	77816	+	0
	>4580745	/2757			
	len =	777	nex =	1	
55	Sngl	11301	12077	+	0
	>4580745	/32177			
60	len =	757	nex =	1	

1450

	Intr	15767	16020	+	0
	Intr	16168	16291	+	0
	Intr	16405	16509	+	0
	Term	16791	17076	+	0
5	>4581084 /34830				
	len =	1601	nex =	3	
10	Term	22349	21931	-	0
	Intr	23000	22548	-	0
	Init	23531	23196	-	0
	>4581084 /28597				
15	len =	1825	nex =	2	
	Term	40647	39196	-	0
	Init	41020	40887	-	0
20	>4581084 /41311				
	len =	2171	nex =	7	
25	Init	41924	42481	+	0
	Intr	42756	42849	+	0
	Intr	42930	42963	+	0
	Intr	43048	43102	+	0
	Intr	43184	43267	+	0
30	Intr	43366	43653	+	0
	Term	43750	44094	+	0
	>4581084 /37650				
35	len =	2337	nex =	8	
	Term	3055	2967	-	0
	Intr	3229	3167	-	0
	Intr	3411	3322	-	0
40	Intr	3664	3604	-	0
	Intr	3843	3749	-	0
	Intr	4154	4065	-	0
	Intr	4317	4231	-	0
	Init	4792	4411	-	0
45	>4581084 /17428				
	len =	2451	nex =	7	
50	Term	3055	2967	-	0
	Intr	3229	3167	-	0
	Intr	3411	3322	-	0
	Intr	3664	3604	-	0
	Intr	3843	3749	-	0
55	Intr	4154	4065	-	0
	Init	4317	4231	-	0
	>4581084 /93110				
60	len =	1289	nex =	3	

1451

	Term	60581	60181	-	0
	Intr	61120	61028	-	0
	Init	61458	61210	-	0
5	>4581084	/38429			
	len =	2273	nex =	6	
10	Term	60581	60221	-	0
	Intr	61120	61028	-	0
	Intr	61458	61210	-	0
	Intr	61709	61557	-	0
15	Intr	62108	61784	-	0
	Init	62493	62191	-	0
	>4581084	/5677			
	len =	943	nex =	2	
20	Init	72815	73073	+	0
	Term	73160	73757	+	0
	>4581103	/36296			
25	len =	1853	nex =	4	
	Init	105971	106313	+	0
	Intr	106543	106838	+	0
30	Intr	107212	107352	+	0
	Term	107448	107823	+	0
	>4581103	/119868			
35	len =	638	nex =	2	
	Term	25527	25372	-	0
	Init	26009	25647	-	0
40	>4581103	/26940			
	len =	1076	nex =	3	
	Init	49009	49124	+	0
45	Intr	49202	49461	+	0
	Term	49704	50084	+	0
	>4581103	/23995			
50	len =	755	nex =	1	
	Sngl	5195	5949	+	0
	>4581103	/31917			
55	len =	2086	nex =	3	
	Init	52571	52873	+	0
	Intr	54100	54217	+	0
60	Term	54298	54656	+	0

1452

>4581103 /27679

```

5      len =      1056      nex =          4
      Init  72192  72336      +      0
      Intr  72438  72565      +      0
      Intr  72751  72855      +      0
      Term  72981  73247      +      0

```

10
>4581103 /40637

	len =	2148	nex =	6	
15	Term	86242	85852	-	0
	Intr	86439	86344	-	0
	Intr	86717	86593	-	0
	Intr	87013	86870	-	0
	Intr	87427	87104	-	0
20	Init	87999	87653	-	0

>4581103 /19417

25	len =	1281	nex =	4	
	Term	88901	88344	-	0
	Intr	89054	88975	-	0
	Intr	89191	89149	-	0
	Init	89329	89270	-	0

>4581103 /23735

```

len =      1579    nex =      2
35  Init  90125    90516    +      0
    Term  90763    91703    +      0

```

>4581138 /17193

40	len =	2079	nex =	6	
	Init	17050	17171	+	0
	Intr	17298	17403	+	0
	Intr	17518	17646	+	0
45	Intr	17927	18071	+	0
	Intr	18203	18760	+	0
	Term	18845	19128	+	0

>4581138 /5234

```

50      len =      812    nex =      2
      Init 18317  18760      +      0
      Term 18845  19128      +      0

```

>4581138 /13024

```

len =      1680    nex =          3
60    Term 46797    45819      -      0

```

1453

	Intr	47110	46883	-	0
	Init	47498	47408	-	0
	>4581138 /18705				
5	len =	550	nex =	1	
	Sngl	52038	51498	-	0
10	>4581138 /39339				
	len =	291	nex =	1	
	Sngl	70179	69889	-	0
15	>4581138 /9198				
	len =	646	nex =	3	
20	Term	70209	69891	-	0
	Intr	70349	70293	-	0
	Init	70536	70438	-	0
	>4581138 /12548				
25	len =	2972	nex =	12	
	Term	70209	69940	-	0
	Intr	70349	70293	-	0
30	Intr	70629	70438	-	0
	Intr	70921	70727	-	0
	Intr	71094	71017	-	0
	Intr	71319	71182	-	0
	Intr	71648	71466	-	0
35	Intr	71860	71780	-	0
	Intr	72035	71949	-	0
	Intr	72187	72140	-	0
	Intr	72413	72285	-	0
	Init	72911	72769	-	0
40	>4581138 /12298				
	len =	3528	nex =	16	
45	Init	88301	88493	+	0
	Intr	88550	88646	+	0
	Intr	88730	88827	+	0
	Intr	88922	89077	+	0
	Intr	89176	89253	+	0
50	Intr	89364	89495	+	0
	Intr	89613	89693	+	0
	Intr	89822	89893	+	0
	Intr	89997	90086	+	0
	Intr	90230	90374	+	0
55	Intr	90498	90586	+	0
	Intr	90705	90800	+	0
	Intr	90897	90964	+	0
	Intr	91053	91152	+	0
	Intr	91480	91539	+	0
60	Term	91677	91828	+	0

2025 RELEASE UNDER E.O. 14176

1454

>4581138 /20959

5	len =	2279	nex =	7	
	Term	92547	92107	-	0
	Intr	92826	92611	-	0
	Intr	93078	92954	-	0
	Intr	93363	93174	-	0
10	Intr	93544	93450	-	0
	Intr	93830	93620	-	0
	Init	94385	94134	-	0

>4581161 /6542

15	len =	974	nex =	2	
	Term	103655	103129	-	0
	Init	104102	104014	-	0

>4581161 /4401

20	len =	944	nex =	2	
25	Term	103655	103178	-	0
	Init	104121	104014	-	0

>4581161 /3929

30	len =	2594	nex =	5	
	Term	101826	101547	-	0
	Intr	102280	102202	-	0
	Intr	102758	102630	-	0
35	Intr	103655	103281	-	0
	Init	104140	104014	-	0

>4581161 /207629

40	len =	1078	nex =	1	
	Sngl	104140	104014	-	0

>4581161 /153194

45	len =	1076	nex =	2	
	Term	103655	103065	-	0
	Init	104140	104014	-	0

>4581161 /113443

50	len =	1014	nex =	1	
55	Sngl	104140	104014	-	0

>4581161 /105261

60	len =	1012	nex =	2	
----	-------	------	-------	---	--

000000-000000

1455

	Term	103655	103129	-	0
	Init	104140	104014	-	0
5	>4581161 /23214				
	len =	970	nex =	2	
	Term	103655	103172	-	0
10	Init	104140	104014	-	0
	>4581161 /92991				
	len =	970	nex =	2	
15	Term	103655	103178	-	0
	Init	104140	104014	-	0
	>4581161 /109952				
20	len =	951	nex =	2	
	Term	103655	103190	-	0
	Init	104140	104014	-	0
25	>4581161 /118540				
	len =	597	nex =	2	
	Term	103655	103544	-	0
30	Init	104140	104014	-	0
	>4581161 /18215				
	len =	2507	nex =	1	
35	Sngl	104144	104014	-	0
	>4581161 /24845				
40	len =	970	nex =	2	
	Term	103655	103177	-	0
	Init	104144	104014	-	0
45	>4581161 /24667				
	len =	970	nex =	2	
	Term	103655	103178	-	0
50	Init	104144	104014	-	0
	>4581161 /3416				
	len =	1815	nex =	5	
55	Init	12372	12490	+	0
	Intr	13293	13348	+	0
	Intr	13421	13473	+	0
	Intr	13699	13817	+	0
60	Term	13921	14186	+	0

1456

	>4581161	/9243		
5	len =	943 nex =	2	
	Init	12408 12490	+	0
	Term	13293 13350	+	0
10	>4581161	/1618		
	len =	698 nex =	2	
	Init	15445 15838	+	0
	Term	15921 16142	+	0
15	>4581161	/6509		
	len =	395 nex =	1	
20	Sngl	80928 80534	-	0
	>4581161	/12929		
25	len =	396 nex =	1	
	Sngl	80929 80534	-	0
	>4581161	/24190		
30	len =	933 nex =	2	
	Term	81003 80534	-	0
	Init	81466 81341	-	0
35	>4581161	/39985		
	len =	934 nex =	2	
	Term	81003 80534	-	0
40	Init	81467 81341	-	0
	>4581161	/20104		
	len =	911 nex =	2	
45	Term	81003 80557	-	0
	Init	81467 81341	-	0
	>4581161	/109026		
50	len =	753 nex =	2	
	Term	81003 80715	-	0
	Init	81467 81341	-	0
55	>4581161	/34326		
	len =	938 nex =	2	
60	Term	81003 80534	-	0

1457

	Init	81471	81341	-	0
	>4581161 /12459				
5	len =	915	nex =	2	
	Term	81003	80557	-	0
	Init	81471	81341	-	0
10	>4581161 /103273				
	len =	910	nex =	0	
	>4581161 /102088				
15	len =	671	nex =	2	
	Term	81003	80801	-	0
	Init	81471	81341	-	0
20	>4582411 /33336				
	len =	2810	nex =	7	
25	Init	45451	45474	+	0
	Intr	45798	45903	+	0
	Intr	46005	46073	+	0
	Intr	46212	46705	+	0
	Intr	47033	47294	+	0
30	Intr	47380	47588	+	0
	Term	47685	47897	+	0
	>4582411 /21223				
35	len =	2816	nex =	7	
	Init	45451	45474	+	0
	Intr	45798	45903	+	0
	Intr	46005	46073	+	0
40	Intr	46212	46705	+	0
	Intr	47033	47294	+	0
	Intr	47380	47588	+	0
	Term	47685	47903	+	0
45	>4582411 /12397				
	len =	1719	nex =	5	
	Init	48106	48233	+	0
50	Intr	48506	48720	+	0
	Intr	49001	49081	+	0
	Intr	49174	49521	+	0
	Term	49615	49824	+	0
55	>4582428 /13146				
	len =	1111	nex =	2	
	Init	14118	14404	+	0
60	Term	14774	15228	+	0

ORIGINAL RELEASED

1458

>4582428 /16737

5	len =	1407	nex =	5	
	Init	18518	18583	+	0
	Intr	18928	19009	+	0
	Intr	19112	19183	+	0
	Intr	19501	19553	+	0
10	Term	19701	19924	+	0

>4582428 /97480

```

15      len =      1161      nex =          4
        Init  18928   19009      +          0
        Intr  19112   19183      +          0
        Intr  19501   19553      +          0
        Term  19701   19911      +          0

```

>4582437 /2900

```

len =      397    nex =      1
25    Snq1  19712  19316      -      0

```

>4582437 /99899

```

30      len =      680      nex =      1
        Sngl  7826      8505      +      0

```

>4582444 /103070

```

35      len =      790      nex =          2
      Init    2508    3031      +          0
      Term    3115    3290      +          0

```

```
40 >4582444 /37180
len = 1440 nex = 1
```

Sng1	25314	26753	+	0
------	-------	-------	---	---

>4582444 / 123564

```

len =      1450    nex =      1
50      Sngl 25977  26038      +      0

```

>4582444 /25550

```

55      len =      1675      nex =          5
      Term 27247 27080 - 0
      Intr 27553 27435 - 0
      Intr 27722 27661 - 0
      Intr 28224 28148 - 0
60      Init 28606 28475 - 0

```

1459

>4582444 /19236

5	len =	651	nex =	1	
	Sngl	34810	35460	+	0

>4582444 /9568

10	len =	1311	nex =	5	
	Init	39087	39157	+	0
	Intr	39241	39388	+	0
	Intr	39469	39640	+	0
15	Intr	39922	40108	+	0
	Term	40198	40397	+	0

>4582444 /15190

20	len =	718	nex =	2	
	Init	39945	40108	+	0
	Term	40198	40662	+	0

25 >4582444 /15416

	len =	1431	nex =	2	
	Term	62850	62359	-	0
30	Init	63789	63368	-	0

>4582444 /74

35	len =	1311	nex =	1	
	Sngl	71021	70539	-	0

>4582444 /25159

40	len =	1106	nex =	3	
	Term	75373	75084	-	0
	Intr	75585	75448	-	0
	Init	76189	76007	-	0

45 >4582444 /94470

	len =	430	nex =	1	
50	Sngl	78400	77979	-	0

>4584339 /13257

	len =	2291	nex =	6	
55	Term	19077	18435	-	0
	Intr	19277	19154	-	0
	Intr	19501	19359	-	0
	Intr	19724	19589	-	0
60	Intr	20156	19825	-	0

1460

	Init	20725	20474	-	0
	>4584339	/641			
5	len =	1150	nex =	1	
	Sngl	26683	25535	-	0
	>4584339	/552			
10	len =	1065	nex =	2	
	Init	28913	29192	+	0
	Term	29589	29977	+	0
15	>4584351	/17506			
	len =	490	nex =	1	
20	Sngl	30155	29673	-	0
	>4584351	/41808			
	len =	521	nex =	1	
25	Sngl	30170	30061	-	0
	>4584351	/143435			
30	len =	471	nex =	1	
	Sngl	30170	29700	-	0
	>4584351	/14272			
35	len =	575	nex =	1	
	Sngl	30249	29675	-	0
40	>4584351	/91878			
	len =	1832	nex =	5	
	Term	30771	30505	-	0
45	Intr	31297	31037	-	0
	Intr	31481	31377	-	0
	Intr	31721	31591	-	0
	Init	32336	32137	-	0
50	>4584387	/7818			
	len =	1390	nex =	1	
	Sngl	19033	19118	+	0
55	>4584387	/12275			
	len =	490	nex =	1	
60	Sngl	31615	31127	-	0

1461

>4584387 /18912

5	len =	5325	nex =	20	
	Init	31825	31912	+	0
	Intr	32018	32165	+	0
	Intr	32301	32493	+	0
	Intr	32573	32662	+	0
10	Intr	32903	32965	+	0
	Intr	33174	33243	+	0
	Intr	33394	33458	+	0
	Intr	33795	33849	+	0
	Intr	33973	34094	+	0
15	Intr	34178	34293	+	0
	Intr	34678	34747	+	0
	Intr	34996	35100	+	0
	Intr	35202	35291	+	0
	Intr	35374	35522	+	0
20	Intr	35706	35817	+	0
	Intr	35900	35983	+	0
	Intr	36106	36186	+	0
	Intr	36275	36499	+	0
	Intr	36584	36710	+	0
25	Term	36857	37134	+	0

>4584387 /6665

30	len =	2598	nex =	7	
	Init	38193	38274	+	0
	Intr	38429	38504	+	0
	Intr	38670	38833	+	0
	Intr	38917	39067	+	0
35	Intr	39199	39321	+	0
	Intr	39439	39534	+	0
	Term	39625	40136	+	0

>4584519 /33414

40	len =	2899	nex =	10	
	Term	97811	97456	-	0
	Intr	98170	97906	-	0
45	Intr	98475	98309	-	0
	Intr	98711	98568	-	0
	Intr	99028	98919	-	0
	Intr	99221	99113	-	0
	Intr	99408	99319	-	0
50	Intr	99531	99493	-	0
	Intr	99800	99606	-	0
	Init	99974	99924	-	0

>4584519 /21149

55	len =	1545	nex =	7	
	Term	100825	100655	-	0
	Intr	101015	100926	-	0
60	Intr	101223	101105	-	0

0050526-7541600

1462

	Intr	101620	101506	-	0
	Intr	101832	101736	-	0
	Intr	101991	101928	-	0
	Init	102199	102148	-	0
5	>4584519 /6308				
	len =	889	nex =	2	
10	Init	28246	28587	+	0
	Term	28683	29134	+	0
	>4584519 /28170				
15	len =	1464	nex =	6	
	Term	34134	34075	-	0
	Intr	34343	34218	-	0
	Intr	34584	34431	-	0
20	Intr	34767	34659	-	0
	Intr	34979	34916	-	0
	Init	35201	35079	-	0
	>4584519 /12980				
25	len =	612	nex =	1	
	Sngl	38892	38281	-	0
30	>4584531 /1259				
	len =	1881	nex =	3	
	Init	2041	2614	+	0
35	Intr	3185	3256	+	0
	Term	3662	3921	+	0
	>4584531 /10016				
40	len =	1471	nex =	4	
	Term	41067	40537	-	0
	Intr	41273	41234	-	0
	Intr	41510	41378	-	0
45	Init	41872	41792	-	0
	>4584531 /42200				
	len =	430	nex =	1	
50	Sngl	56875	56451	-	0
	>4584531 /38181				
55	len =	3698	nex =	4	
	Term	4219	3881	-	0
	Intr	4842	4704	-	0
	Intr	5250	5029	-	0
60	Init	7571	7406	-	0

2025 RELEASE UNDER E.O. 14176

1463

	>4584841	/37069		
5	len =	1301	nex =	1
	Sngl	1552	252	- 0
	>4584841	/16143		
10	len =	1050	nex =	2
	Term	75412	74936	- 0
	Init	75985	75751	- 0
15	>4584841	/159247		
	len =	236	nex =	1
20	Sngl	75985	75750	- 0
	>4584841	/6145		
	len =	910	nex =	2
25	Term	75412	75090	- 0
	Init	75994	75751	- 0
	>4584841	/123475		
30	len =	732	nex =	2
	Term	75412	75265	- 0
	Init	75996	75751	- 0
35	>4584841	/9874		
	len =	1765	nex =	5
40	Init	81507	81619	+ 0
	Intr	81801	81961	+ 0
	Intr	82057	82137	+ 0
	Intr	82622	82745	+ 0
	Term	82877	83265	+ 0
45	>4585890	/120166		
	len =	653	nex =	1
50	Sngl	12573	11926	- 0
	>4585890	/13800		
	len =	2230	nex =	7
55	Term	19173	18816	- 0
	Intr	19409	19311	- 0
	Intr	19822	19718	- 0
	Intr	20057	19911	- 0
	Intr	20227	20141	- 0
60	Intr	20363	20307	- 0

1464

	Init	21042	20494	-	0
	>4585890 /23237				
5	len =	1936	nex =	8	
	Term	21526	21333	-	0
	Intr	21725	21662	-	0
	Intr	21895	21830	-	0
10	Intr	22074	22013	-	0
	Intr	22419	22332	-	0
	Intr	22565	22505	-	0
	Intr	22733	22650	-	0
	Init	23268	22886	-	0
15	>4585890 /101259				
	len =	2530	nex =	9	
20	Init	27106	27159	+	0
	Intr	27272	27335	+	0
	Intr	27474	27530	+	0
	Intr	27682	27788	+	0
	Intr	27891	27929	+	0
25	Intr	28113	28201	+	0
	Intr	28292	28370	+	0
	Intr	28459	28506	+	0
	Term	28596	28976	+	0
30	>4585890 /6787				
	len =	1070	nex =	1	
	Sngl	35922	34853	-	0
35	>4585890 /6066				
	len =	1954	nex =	7	
40	Init	63417	63643	+	0
	Intr	63802	63857	+	0
	Intr	63959	64073	+	0
	Intr	64155	64265	+	0
	Intr	64360	64548	+	0
45	Intr	64929	64985	+	0
	Term	65108	65370	+	0
	>4585891 /33396				
50	len =	1691	nex =	3	
	Term	13026	12969	-	0
	Intr	13198	13099	-	0
	Init	13882	13459	-	0
55	>4585891 /99687				
	len =	1230	nex =	1	
60	Sngl	26998	27650	+	0

1465

>4585891 /38876

5	len =	2154	nex =	8	
	Term	57147	56769	-	0
	Intr	57348	57237	-	0
	Intr	57473	57439	-	0
	Intr	57784	57707	-	0
10	Intr	58148	58008	-	0
	Intr	58290	58225	-	0
	Intr	58533	58389	-	0
	Init	58922	58608	-	0

15 >4585891 /108157

	len =	407	nex =	1	
	Sngl	6299	5893	-	0

20 >4585891 /113049

	len =	1133	nex =	2	
25	Init	67122	67207	+	0
	Term	67928	68110	+	0

>4585891 /3006

30	len =	2536	nex =	10	
	Term	68750	68545	-	0
	Intr	68912	68844	-	0
	Intr	69081	69011	-	0
35	Intr	69306	69228	-	0
	Intr	69556	69392	-	0
	Intr	69684	69640	-	0
	Intr	69849	69775	-	0
	Intr	70301	70125	-	0
40	Intr	70649	70386	-	0
	Init	71080	70870	-	0

>4585891 /22546

45	len =	1469	nex =	1	
	Sngl	75226	73758	-	0

>4585891 /16254

50	len =	1456	nex =	1	
	Sngl	75226	73771	-	0

55 >4585891 /39762

	len =	1633	nex =	3	
	Term	6306	5906	-	0
60	Intr	6905	6416	-	0

2025-09-26 10:00:00

09536-0710

— 0

60

1467

>4585952 /857

len = 474 nex = 2

5	Term	33620	33385	-	0
	Init	33853	33807	-	0

>4585952 /23092

10 len = 568 nex = 1

Sngl 39340 39907 + 0

>4585952 /205610

15 len = 550 nex = 1

Sngl 39349 39891 + 0

>4585952 /15789

len = 430 nex = 1

25 Sngl 39378 39802 + 0

>4585952 /11306

len = 3130 nex = 12

30	Term	45205	44960	-	0
	Intr	45443	45287	-	0
	Intr	45633	45542	-	0
	Intr	45776	45712	-	0
	Intr	45925	45860	-	0
35	Intr	46136	46005	-	0
	Intr	46398	46225	-	0
	Intr	46643	46485	-	0
	Intr	46814	46727	-	0
	Intr	46979	46898	-	0
40	Intr	47184	47067	-	0
	Init	47333	47268	-	0

>4585952 /20491

45 len = 2610 nex = 7

	Term	60891	60337	-	0
	Intr	61479	61333	-	0
	Intr	61646	61593	-	0
50	Intr	61921	61751	-	0
	Intr	62418	62209	-	0
	Intr	62609	62523	-	0
	Init	62935	62696	-	0

55 >4585952 /1496

len = 610 nex = 1

60 Sngl 91280 91884 + 0

000100-9252520

1468

>4586024 /207026

len = 970 nex = 2

5	Term	24258	24031	-	0
	Init	24993	24849	-	0

>4586024 /39169

10 len = 1138 nex = 2

	Init	25959	26001	+	0
	Term	26086	26319	+	0

15 >4586024 /24343

len = 768 nex = 2

20	Term	32828	32380	-	0
	Init	33147	32916	-	0

>4586024 /6387

25 len = 2010 nex = 5

	Init	35746	35924	+	0
	Intr	36211	36327	+	0
	Intr	36439	36609	+	0
	Intr	36707	36919	+	0
30	Term	37053	37444	+	0

>4586024 /6176

35 len = 1690 nex = 5

	Init	35746	35924	+	0
	Intr	36211	36327	+	0
	Intr	36439	36609	+	0
	Intr	36707	36919	+	0
40	Term	37053	37417	+	0

>4586024 /25528

45 len = 1217 nex = 2

	Term	42248	41624	-	0
	Init	42832	42420	-	0

>4586024 /15198

50 len = 1372 nex = 5

	Init	7326	7385	+	0
	Intr	7539	7640	+	0
55	Intr	7863	7992	+	0
	Intr	8093	8205	+	0
	Term	8475	8697	+	0

>4586065 /41879

60

2025 RELEASE UNDER E.O. 14176

1469

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

	len =	745	nex =	1	
	Sngl	52682	51938	-	0
5	>4586098	/21839			
	len =	2814	nex =	10	
10	Term	8727	8320	-	0
	Intr	8910	8813	-	0
	Intr	9254	9144	-	0
	Intr	9412	9361	-	0
	Intr	9590	9514	-	0
	Intr	9773	9717	-	0
15	Intr	9978	9878	-	0
	Intr	10163	10067	-	0
	Intr	10363	10251	-	0
	Init	11133	10862	-	0
20	>4586098	/2882			
	len =	1731	nex =	4	
	Term	27031	26717	-	0
25	Intr	27342	27125	-	0
	Intr	27658	27422	-	0
	Init	28447	27982	-	0
	>4586098	/45			
30	len =	1285	nex =	1	
	Sngl	48639	49307	+	0
35	>4586098	/32461			
	len =	1772	nex =	5	
	Term	61744	61488	-	0
40	Intr	62495	62316	-	0
	Intr	62716	62607	-	0
	Intr	62865	62827	-	0
	Init	63259	62978	-	0
45	>4586098	/34781			
	len =	2678	nex =	5	
	Init	65140	65679	+	0
50	Intr	66213	66365	+	0
	Intr	66468	66604	+	0
	Intr	66694	66921	+	0
	Term	67078	67817	+	0
55	>4586098	/15574			
	len =	2030	nex =	5	
	Init	72537	73258	+	0
60	Intr	73723	73790	+	0

1470

Intr	73883	74014	+	0
Intr	74091	74252	+	0
Term	74331	74555	+	0

5 >4586098 /35110

len = 2352 nex = 5

10	Init	87367	87850	+	0
	Intr	88069	88391	+	0
	Intr	88884	89031	+	0
	Intr	89142	89174	+	0
	Term	89266	89718	+	0

15 >4586241 /14095

len = 1133 nex = 1

Sngl 11254 12386 + 0

20 >4586241 /31570

len = 2376 nex = 8

25	Init	18202	18375	+	0
	Intr	18521	18586	+	0
	Intr	18671	18811	+	0
	Intr	18884	18961	+	0
	Intr	19370	19420	+	0
30	Intr	19523	19557	+	0
	Intr	19641	19752	+	0
	Term	19852	20232	+	0

35 >4586241 /36486

len = 1427 nex = 0

>4586241 /26799

40 len = 474 nex = 1

Sngl 24750 24277 - 0

45 >4586241 /12532

len = 2237 nex = 10

	Term	47904	47677	-	0
	Intr	48038	47983	-	0
50	Intr	48211	48118	-	0
	Intr	48426	48296	-	0
	Intr	48622	48519	-	0
	Intr	48849	48764	-	0
	Intr	49074	49005	-	0
55	Intr	49267	49159	-	0
	Intr	49471	49364	-	0
	Init	49913	49842	-	0

60 >4586241 /34400

1471

	len =	1308	nex =	2	
	Init	51838	52152	+	0
	Term	52518	53145	+	0
5	>4586241	/37484			
	len =	4542	nex =	2	
	Init	72812	73062	+	0
	Term	77290	77353	+	0
10	>4586241	/100613			
	len =	4760	nex =	3	
	Init	72812	73062	+	0
	Intr	77290	77355	+	0
	Term	77455	77571	+	0
15	>4586241	/516			
	len =	5258	nex =	5	
	Init	72812	73062	+	0
	Intr	77290	77355	+	0
	Intr	77455	77577	+	0
	Intr	77692	77784	+	0
	Term	77856	78069	+	0
20	>4586241	/41510			
	len =	910	nex =	2	
	Term	75339	74906	-	0
	Init	75807	75569	-	0
25	>4586241	/35907			
	len =	1110	nex =	5	
	Init	76894	77206	+	0
	Intr	77290	77355	+	0
	Intr	77455	77577	+	0
	Intr	77692	77784	+	0
	Term	77856	78003	+	0
30	>4586241	/17520			
	len =	790	nex =	2	
	Init	8059	8299	+	0
	Term	8495	8848	+	0
35	>4587582	/125961			
	len =	490	nex =	1	
	Sngl	13888	13411	-	0
40					
45					
50					
55					
60					

1472

>4587582 /31554

```
len = 1930    nex = 4
```

5	Term	21991	21883	-	0
	Intr	22243	22097	-	0
	Intr	22707	22622	-	0
	Init	23280	23003	-	0

10 >4587582 /4595

```
len = 1353    nex = 1
```

Snq1	23593	24945	+	0
------	-------	-------	---	---

>4587582 / 803

```
len = 2112    nex = 2
```

20	Init	26649	27735	+	0
	Term	28584	28760	+	0

>4587582 /16530

```

25      len =      1096    nex =          2
      Term 29562    29488      -          0
      Init 30202    30172      -          0

```

30 >4587582 /103045

```
len = 1270    nex = 2
```

	Term	29562	29325	-	0
35	Init	30202	30172	-	0

>4587582 /10073

```
len = 1665    nex = 2
```

Term	29562	28975	-	0
Init	30202	30172	-	0

>4587641 /19234

```
len =      749    nex =      1
```

Snq1	129286	128538	-	0
------	--------	--------	---	---

50 >4587641 /35921

```
len = 2548    nex = 5
```

55	Init	134987	135707	+	0
	Intr	135938	136040	+	0
	Intr	136602	136657	+	0
	Intr	136808	136972	+	0
	Term	137280	137534	+	0

60 >4587641 /29173

1473

	len =	1616	nex =	4	
	Init	26708	26786	+	0
5	Intr	27042	27199	+	0
	Intr	27523	27614	+	0
	Term	27722	28051	+	0
	>4587641 /20583				
10	len =	1234	nex =	4	
	Init	26708	26786	+	0
	Intr	27042	27199	+	0
15	Intr	27523	27614	+	0
	Term	27722	27934	+	0
	>4587641 /3997				
20	len =	2393	nex =	4	
	Init	31992	32060	+	0
	Intr	32147	32250	+	0
	Intr	32360	32790	+	0
25	Term	32881	33710	+	0
	>4587641 /9992				
	len =	1188	nex =	4	
30	Init	57159	57189	+	0
	Intr	57295	57437	+	0
	Intr	57696	57818	+	0
	Term	57908	58346	+	0
35	>4587641 /41809				
	len =	1062	nex =	3	
40	Init	60850	61072	+	0
	Intr	61195	61445	+	0
	Term	61534	61911	+	0
	>4587641 /35151				
45	len =	649	nex =	2	
	Term	81448	81337	-	0
	Init	81985	81802	-	0
50	>4587641 /12738				
	len =	359	nex =	1	
55	Sngl	82158	81800	-	0
	>4587641 /7074				
60	len =	690	nex =	1	

1474

	Sngl	82161	81802	-	0
	>4587641	/108142			
5	len =	831	nex =	2	
	Term	81448	81334	-	0
	Init	82164	81802	-	0
10	>4587677	/42117			
	len =	1397	nex =	4	
	Term	4964	4716	-	0
15	Intr	5403	5201	-	0
	Intr	5650	5565	-	0
	Init	6112	5814	-	0
	>4587986	/14922			
20	len =	561	nex =	1	
	Sngl	19040	19589	+	0
25	>4587986	/6782			
	len =	400	nex =	1	
	Sngl	37458	37260	-	0
30	>4589409	/6760			
	len =	715	nex =	2	
35	Init	15185	15608	+	0
	Term	15705	15899	+	0
	>4589409	/15004			
40	len =	730	nex =	1	
	Sngl	22595	21875	-	0
	>4589409	/29063			
45	len =	1952	nex =	4	
	Term	23209	22949	-	0
	Intr	23499	23305	-	0
50	Intr	23954	23842	-	0
	Init	24900	24844	-	0
	>4589409	/16621			
55	len =	2096	nex =	5	
	Init	25892	26032	+	0
	Intr	26337	26445	+	0
	Intr	26520	26581	+	0
60	Intr	26672	26775	+	0

1475

	Term	26873	27227	+	0
	>4589409	/28003			
5	len =	1663	nex =	2	
	Init	33218	33581	+	0
	Term	34356	34880	+	0
10	>4589409	/32842			
	len =	1305	nex =	1	
	Sngl	36996	38300	+	0
15	>4589410	/41421			
	len =	1995	nex =	4	
20	Init	13570	13637	+	0
	Intr	13729	14122	+	0
	Intr	14271	14884	+	0
	Term	14990	15286	+	0
25	>4589410	/6022			
	len =	1463	nex =	3	
30	Init	13320	13637	+	0
	Intr	13729	14122	+	0
	Term	14271	14782	+	0
	>4589410	/31579			
35	len =	3250	nex =	10	
	Init	25051	25297	+	0
	Intr	25751	25844	+	0
	Intr	25940	26056	+	0
40	Intr	26147	26219	+	0
	Intr	26327	26445	+	0
	Intr	26548	26669	+	0
	Intr	26906	27082	+	0
	Intr	27194	27338	+	0
45	Intr	27791	27669	+	0
	Term	27843	28293	+	0
	>4589410	/32783			
50	len =	716	nex =	3	
	Init	32322	32516	+	0
	Intr	32607	32735	+	0
	Term	32829	32876	+	0
55	>4589410	/22152			
	len =	1661	nex =	5	
60	Term	40659	40201	-	0

1476

	Intr	40967	40894	-	0
	Intr	41125	41044	-	0
	Intr	41359	41262	-	0
	Init	41861	41719	-	0
5	>4589410 /38797				
	len =	1392	nex =	1	
10	Sngl	49398	50789	+	0
	>4589410 /157709				
	len =	623	nex =	1	
15	Sngl	53543	54165	+	0
	>4589410 /21863				
20	len =	1341	nex =	1	
	Sngl	63197	61857	-	0
	>4589410 /41153				
25	len =	1351	nex =	2	
	Init	67494	68408	+	0
	Term	68630	68844	+	0
30	>4589411 /2612				
	len =	795	nex =	1	
35	Sngl	11554	11467	-	0
	>4589411 /12760				
	len =	1958	nex =	6	
40	Term	11151	10758	-	0
	Intr	11356	11253	-	0
	Intr	11566	11467	-	0
	Intr	12211	12152	-	0
45	Intr	12387	12316	-	0
	Init	12562	12491	-	0
	>4589411 /7631				
50	len =	2679	nex =	9	
	Term	11151	10921	-	0
	Intr	11356	11253	-	0
	Intr	11566	11467	-	0
55	Intr	12211	12152	-	0
	Intr	12387	12316	-	0
	Intr	12562	12491	-	0
	Intr	12737	12645	-	0
	Intr	12928	12861	-	0
60	Init	13599	13285	-	0

2025 RELEASE UNDER E.O. 14176

1477

>4589412 /144945

5	len =	1238	nex =	4	
	Init	13026	13085	+	0
	Intr	13230	13310	+	0
	Intr	13425	13664	+	0
	Term	13744	14126	+	0

10

>4589412 /25062

	len =	2039	nex =	3	
15	Term	14762	14637	-	0
	Intr	15769	15710	-	0
	Init	16311	15854	-	0

20

>4589412 /22430

	len =	4909	nex =	15	
	Init	17362	17458	+	0
	Intr	17622	17795	+	0
25	Intr	17885	17936	+	0
	Intr	18026	18133	+	0
	Intr	18262	18346	+	0
	Intr	18636	18690	+	0
	Intr	18783	19226	+	0
30	Intr	19904	19963	+	0
	Intr	20194	20292	+	0
	Intr	20386	20496	+	0
	Intr	20651	20722	+	0
	Intr	20815	20931	+	0
35	Intr	21132	21281	+	0
	Intr	21415	21515	+	0
	Term	21862	22270	+	0

40

>4589412 /92349

	len =	1260	nex =	4	
	Term	1333	917	-	0
	Intr	1647	1400	-	0
45	Intr	2010	1852	-	0
	Init	2176	2133	-	0

>4589412 /6550

50	len =	940	nex =	3	
	Term	35193	34956	-	0
	Intr	35603	35471	-	0
	Init	35895	35698	-	0

55

>4589412 /91769

	len =	1450	nex =	5	
60	Term	35193	34965	-	0

1478

	Intr	35603	35471	-	0
	Intr	35977	35698	-	0
	Intr	36224	36176	-	0
	Init	36412	36339	-	0
5	>4589412 /37702				
	len =	2953	nex =	8	
10	Term	38862	38499	-	0
	Intr	39017	38958	-	0
	Intr	39247	39138	-	0
	Intr	39446	39339	-	0
	Intr	39725	39558	-	0
15	Intr	39942	39819	-	0
	Intr	40327	40270	-	0
	Init	41451	41110	-	0
	>4589412 /19542				
20	len =	859	nex =	3	
	Init	63381	63483	+	0
	Intr	63632	63707	+	0
25	Term	63802	64239	+	0
	>4589412 /26264				
	len =	600	nex =	1	
30	Sngl	70812	70213	-	0
	>4589414 /92314				
35	len =	790	nex =	1	
	Sngl	27496	26709	-	0
	>4589414 /21068				
40	len =	860	nex =	1	
	Sngl	54424	55283	+	0
45	>4589414 /41648				
	len =	2894	nex =	10	
	Init	80194	80450	+	0
50	Intr	80547	80661	+	0
	Intr	80904	80984	+	0
	Intr	81081	81169	+	0
	Intr	81271	81362	+	0
	Intr	81484	81565	+	0
55	Intr	81673	81775	+	0
	Intr	81860	81931	+	0
	Intr	82124	82209	+	0
	Term	82304	83087	+	0
60	>4589415 /4507				

2025-06-16 10:02:00

1479

	len =	855	nex =	2	
	Init	1	35	+	0
5	Term	299	855	+	0
>4589415 /117537					
	len =	2514	nex =	7	
10	Term	12799	12501	-	0
	Intr	12999	12887	-	0
	Intr	13168	13088	-	0
	Intr	13342	13290	-	0
15	Intr	13544	13436	-	0
	Intr	13699	13640	-	0
	Init	14105	14023	-	0
>4589415 /159318					
20	len =	1296	nex =	3	
	Term	17311	16832	-	0
	Intr	17535	17398	-	0
25	Init	18127	17694	-	0
>4589415 /36400					
	len =	1837	nex =	6	
30	Init	18863	19024	+	0
	Intr	19422	19565	+	0
	Intr	19664	19755	+	0
	Intr	19856	20042	+	0
35	Intr	20135	20290	+	0
	Term	20375	20699	+	0
>4589418 /12825					
40	len =	1367	nex =	4	
	Term	91	1	-	0
	Intr	499	185	-	0
	Intr	1228	970	-	0
45	Init	1367	1302	-	0
>4589418 /13931					
	len =	2006	nex =	9	
50	Term	12022	11700	-	0
	Intr	12272	12099	-	0
	Intr	12470	12369	-	0
	Intr	12618	12554	-	0
55	Intr	12784	12701	-	0
	Intr	12976	12904	-	0
	Intr	13171	13070	-	0
	Intr	13490	13371	-	0
	Init	13705	13592	-	0
60					

1480

>4589418 /8249

len = 3294 nex = 13

5	Init	21324	21424	+	0
	Intr	21529	21612	+	0
	Intr	21826	21872	+	0
	Intr	22163	22237	+	0
	Intr	22318	22395	+	0
10	Intr	22608	22670	+	0
	Intr	22801	22871	+	0
	Intr	22975	23074	+	0
	Intr	23213	23317	+	0
	Intr	23570	23632	+	0
15	Intr	23717	23839	+	0
	Intr	23953	24105	+	0
	Term	24189	24497	+	0

>4589419 /24081

20	len =	550	nex =	1	
	Sngl	32416	32957	+	0

>4589419 /106001

25	len =	790	nex =	3	
	Init	32416	32802	+	0
30	Intr	32996	33100	+	0
	Term	33137	33197	+	0

>4589419 /24665

35	len =	730	nex =	1	
	Sngl	35180	34460	-	0

>4589419 /21922

40	len =	2840	nex =	6	
	Term	35177	34454	-	0
	Intr	35520	35348	-	0
45	Intr	35757	35622	-	0
	Intr	36338	35874	-	0
	Intr	36722	36412	-	0
	Init	37293	36801	-	0

>4589419 /10037

	len =	679	nex =	2	
	Term	42723	42430	-	0
55	Init	43108	42926	-	0

>4589421 /28511

60	len =	984	nex =	5	
----	-------	-----	-------	---	--

2025-06-06 14:53:53

1481

	Term	9429	9171	-	0
	Intr	9603	9520	-	0
	Intr	9780	9688	-	0
	Intr	9922	9850	-	0
5	Init	10144	10006	-	0

>4589421 /32718

10	len =	1303	nex =	5	
	Term	25886	25513	-	0
	Intr	26219	26136	-	0
	Intr	26411	26319	-	0
	Intr	26577	26505	-	0
15	Init	26815	26651	-	0

>4589421 /5240

```

20      len =      641      nex =      1
      snq1 61499 62139      +      0

```

>4589421 /20850

```

25      len =      2056      nex =          3
      Term 61249 60746      -      0
      Intr 61952 61661      -      0
      Init 62801 62305      -      0

```

>4589421 /11329

```

len =      442    nex =      1
35      Snql 80041  79600      -      0

```

>4589423 / 33833

```

40      len =      850    nex =      2
      Term  14120    13771    -      0
      Init  14612    14518    -      0

```

>4589423 /36859

```
len =      854  nex =      2
Term 14120 13819 -      0
Init 14672 14518 -      0
```

>4589425 /113229

```

len =      511    nex =      1
55      Snq1 33439 32996      -      0

```

>4589427 /1076

```
len = 498 nex = 1
```

1482

	Sngl	25047	24550	-	0
	>4589427		/17208		
5	len =	1482	nex =	4	
	Term	25597	25201	-	0
	Intr	25959	25691	-	0
	Intr	26181	26067	-	0
10	Init	26540	26433	-	0
	>4589428		/33021		
15	len =	3085	nex =	11	
	Init	11298	11398	+	0
	Intr	12000	12227	+	0
	Intr	12319	12366	+	0
	Intr	12551	12627	+	0
20	Intr	12718	12840	+	0
	Intr	13160	13264	+	0
	Intr	13353	13425	+	0
	Intr	13510	13581	+	0
	Intr	13702	13894	+	0
25	Intr	13978	14107	+	0
	Term	14194	14382	+	0
	>4589428		/39126		
30	len =	2513	nex =	10	
	Init	12000	12227	+	0
	Intr	12319	12366	+	0
	Intr	12551	12627	+	0
35	Intr	12718	12840	+	0
	Intr	13160	13264	+	0
	Intr	13353	13425	+	0
	Intr	13510	13581	+	0
	Intr	13702	13894	+	0
40	Intr	13978	14107	+	0
	Term	14194	14381	+	0
	>4589428		/38630		
45	len =	1819	nex =	4	
	Init	27537	27623	+	0
	Intr	28018	28109	+	0
	Intr	28242	28316	+	0
50	Term	28403	29075	+	0
	>4589428		/21695		
55	len =	2374	nex =	7	
	Init	2805	3043	+	0
	Intr	3162	3288	+	0
	Intr	3751	3868	+	0
	Intr	3992	4128	+	0
60	Intr	4215	4316	+	0

1483

Intr	4412	4504	+	0
Term	4600	5178	+	0

>4589428 /1326

```
5      len =      1476      nex =          3
```

Init	37323	37772	+	0
Intr	38054	38269	+	0

10	Term	38350	38798	+	0
----	------	-------	-------	---	---

>4589428 /28772

```
len = 1258    nex = 4
```

Term	39781	39356	-	0
------	-------	-------	---	---

Intr	39944	39891	-	0
------	-------	-------	---	---

Intr	40129	40040	-	0
------	-------	-------	---	---

Init	40613	40330	-	0
------	-------	-------	---	---

20
 >4589428 /22839

```
len =      496    nex =      1
```

25	Snq1	6020	5525	-	0
----	------	------	------	---	---

>4589430 /37580

```
len = 134    nex = 1
```

Snq1 20755 20888 + 0

>4589430 /35379

```
35      len =      1546      nex =          5
```

Init	36633	36700	+	0
------	-------	-------	---	---

Intr	36804	37042	+	0
------	-------	-------	---	---

Intr	37150	37290	+	0
------	-------	-------	---	---

40	Intr	37379	37813	+	0
----	------	-------	-------	---	---

Term	37902	38175	+	0
------	-------	-------	---	---

>4589430 /26766

```
45      len =      430      nex =      1
```

Snq1	37902	38175	+	0
------	-------	-------	---	---

>4589430 /4357

```

50      len =      2437      nex =      5

```

Init	57993	58176	+	0
------	-------	-------	---	---

Intr	58292	58405	+	0
------	-------	-------	---	---

55	Intr	59163	59390	+	0
----	------	-------	-------	---	---

Intr	59478	59917	+	0
------	-------	-------	---	---

Term	60008	60429	+	0
------	-------	-------	---	---

>4589430 /37065

60

1484

	len =	1885	nex =	4	
	Init	60646	60889	+	0
	Intr	61155	61335	+	0
5	Intr	61583	61724	+	0
	Term	61827	61896	+	0
	>4589430 /110428				
10	len =	535	nex =	2	
	Term	72697	72339	-	0
	Init	72873	72781	-	0
15	>4589432 /3332				
	len =	771	nex =	2	
	Term	15502	15080	-	0
20	Init	15850	15746	-	0
	>4589432 /123678				
25	len =	1078	nex =	2	
	Init	21735	22234	+	0
	Term	22314	22812	+	0
	>4589432 /40179				
30	len =	2617	nex =	5	
	Term	43904	43354	-	0
	Intr	44457	44374	-	0
35	Intr	44853	44548	-	0
	Intr	45547	45471	-	0
	Init	45970	45638	-	0
	>4589433 /12315				
40	len =	194	nex =	1	
	Sngl	10614	10421	-	0
45	>4589434 /28462				
	len =	1771	nex =	2	
	Init	10665	10924	+	0
50	Term	11022	11140	+	0
	>4589434 /104779				
	len =	1639	nex =	2	
55	Init	10667	10924	+	0
	Term	11022	11140	+	0
60	>4589434 /108581				

1485

	len =	1690	nex =	3	
	Init	10667	10924	+	0
	Intr	11022	11140	+	0
5	Term	11602	11648	+	0
	>4589434 /33570				
	len =	1638	nex =	2	
10	Init	10684	10924	+	0
	Term	11022	11140	+	0
	>4589434 /9336				
15	len =	599	nex =	1	
	Sngl	1164	566	-	0
20	>4589434 /39717				
	len =	1792	nex =	5	
	Term	31362	31004	-	0
25	Intr	31686	31448	-	0
	Intr	32066	31860	-	0
	Intr	32345	32156	-	0
	Init	32795	32494	-	0
30	>4589434 /6827				
	len =	2290	nex =	1	
	Sngl	37865	37353	-	0
35	>4589434 /7571				
	len =	2455	nex =	8	
40	Term	43285	43064	-	0
	Intr	43481	43379	-	0
	Intr	43949	43881	-	0
	Intr	44126	44067	-	0
	Intr	44318	44225	-	0
45	Intr	44720	44603	-	0
	Intr	44970	44809	-	0
	Init	45518	45143	-	0
	>4589434 /31680				
50	len =	550	nex =	1	
	Sngl	46938	46393	-	0
55	>4589434 /12344				
	len =	2121	nex =	4	
	Term	52868	52559	-	0
60	Intr	54223	54128	-	0

1486

	Intr	54396	54310	-	0
	Init	54679	54472	-	0
5	>4589434	/15457			
	len =	626	nex =	1	
	Sngl	61850	61225	-	0
10	>4589434	/1876			
	len =	2436	nex =	4	
15	Term	62504	62245	-	0
	Intr	63661	62602	-	0
	Intr	64223	64000	-	0
	Init	64396	64311	-	0
20	>4589434	/159403			
	len =	850	nex =	2	
	Term	67516	66998	-	0
	Init	67838	67591	-	0
25	>4589434	/15623			
	len =	1390	nex =	2	
30	Term	76523	76081	-	0
	Init	77465	76628	-	0
	>4589435	/24826			
35	len =	610	nex =	1	
	Sngl	21600	20996	-	0
40	>4589435	/6424			
	len =	2470	nex =	10	
	Init	29316	29493	+	0
	Intr	29595	29735	+	0
45	Intr	29824	29872	+	0
	Intr	30020	30037	+	0
	Intr	30228	30343	+	0
	Intr	30437	30550	+	0
	Intr	30867	30955	+	0
50	Intr	31183	31260	+	0
	Intr	31353	31460	+	0
	Term	31549	31779	+	0
	>4589435	/10035			
55	len =	1247	nex =	4	
	Init	37688	37775	+	0
	Intr	37871	38014	+	0
60	Intr	38315	38407	+	0

1487

	Term	38519	38934	+	0
	>4589435	/23983			
5	len =	730	nex =	1	
	Sngl	42012	42734	+	0
	>4589435	/22609			
10	len =	459	nex =	1	
	Sngl	69525	69067	-	0
15	>4589435	/93294			
	len =	1273	nex =	5	
20	Term	70107	69919	-	0
	Intr	70459	70342	-	0
	Intr	70608	70552	-	0
	Intr	70794	70727	-	0
	Init	71191	70923	-	0
25	>4589436	/40348			
	len =	1614	nex =	4	
30	Term	16780	16494	-	0
	Intr	17196	17147	-	0
	Intr	17517	17283	-	0
	Init	18107	17946	-	0
35	>4589436	/6991			
	len =	3085	nex =	6	
	Init	29559	29939	+	0
	Intr	29982	30049	+	0
40	Intr	30155	30401	+	0
	Intr	30498	30792	+	0
	Intr	30872	31246	+	0
	Term	31592	32643	+	0
45	>4589436	/2373			
	len =	769	nex =	1	
	Sngl	31994	32762	+	0
50	>4589436	/34946			
	len =	2208	nex =	9	
55	Init	38601	38660	+	0
	Intr	38850	38896	+	0
	Intr	39015	39108	+	0
	Intr	39199	39346	+	0
	Intr	39507	39719	+	0
60	Intr	39820	39954	+	0

2025 RELEASE UNDER E.O. 14176

1488

	Intr	40047	40181	+	0
	Intr	40269	40484	+	0
	Term	40687	40808	+	0
5	>4589437	/42922			
	len =	1778	nex =	2	
10	Term	10942	10397	-	0
	Init	11556	11034	-	0
	>4589437	/92639			
15	len =	610	nex =	1	
	Sngl	55996	55836	-	0
	>4589437	/20558			
20	len =	2456	nex =	5	
	Term	55692	55343	-	0
	Intr	55970	55836	-	0
	Intr	56286	56230	-	0
25	Intr	56657	56523	-	0
	Init	57335	57057	-	0
	>4589437	/108568			
30	len =	3071	nex =	7	
	Init	59166	59592	+	0
	Intr	59799	60132	+	0
	Intr	60223	60263	+	0
35	Intr	60986	61027	+	0
	Intr	61230	61351	+	0
	Intr	61440	61548	+	0
	Term	62046	62236	+	0
40	>4589437	/34275			
	len =	2455	nex =	0	
	>4589437	/14275			
45	len =	840	nex =	1	
	Sngl	8228	7573	-	0
50	>4589437	/41112			
	len =	1440	nex =	7	
55	Init	8889	9041	+	0
	Intr	9162	9359	+	0
	Intr	9462	9551	+	0
	Intr	9629	9718	+	0
	Intr	9803	9883	+	0
	Intr	10024	10069	+	0
60	Term	10176	10328	+	0

2025.06.16 14:56:00

1489

	>4589438	/25545			
5	len =	1215	nex =	2	
	Term	24727	24486	-	0
	Init	25352	25278	-	0
	>4589438	/30064			
10	len =	1570	nex =	5	
	Init	29309	29481	+	0
	Intr	29563	29748	+	0
15	Intr	29831	29993	+	0
	Intr	30080	30266	+	0
	Term	30349	30622	+	0
	>4589438	/21843			
20	len =	989	nex =	1	
	Sngl	46380	47368	+	0
25	>4589438	/22434			
	len =	1518	nex =	4	
	Init	5568	5639	+	0
30	Intr	5947	6024	+	0
	Intr	6128	6406	+	0
	Term	6506	7066	+	0
	>4589439	/10004			
35	len =	463	nex =	1	
	Sngl	40998	41460	+	0
40	>4589439	/26026			
	len =	670	nex =	1	
	Sngl	46310	45649	-	0
45	>4589439	/41488			
	len =	970	nex =	2	
50	Init	47658	48053	+	0
	Term	48141	48323	+	0
	>4589439	/23276			
55	len =	1822	nex =	6	
	Term	63167	62825	-	0
	Intr	63423	63256	-	0
	Intr	63691	63515	-	0
60	Intr	63954	63873	-	0

2025 RELEASE UNDER E.O. 14176

				1490
	Intr	64223	64046	- 0
	Init	64646	64414	- 0
	>4589439	/25458		
5	len =	1950	nex =	5
	Term	65270	64949	- 0
	Intr	65611	65475	- 0
10	Intr	65803	65700	- 0
	Intr	66294	66199	- 0
	Init	66470	66385	- 0
	>4589439	/9482		
15	len =	1129	nex =	1
	Sngl	67879	69007	+ 0
20	>4589439	/42187		
	len =	1190	nex =	1
	Sngl	67901	69090	+ 0
25	>4589440	/30989		
	len =	1881	nex =	5
30	Init	1357	1831	+ 0
	Intr	1938	2061	+ 0
	Intr	2142	2278	+ 0
	Intr	2717	2800	+ 0
	Term	2908	3237	+ 0
35	>4589440	/114071		
	len =	537	nex =	0
40	>4589443	/35696		
	len =	2052	nex =	8
	Init	10117	10334	+ 0
45	Intr	10710	10812	+ 0
	Intr	10911	11018	+ 0
	Intr	11125	11234	+ 0
	Intr	11323	11430	+ 0
	Intr	11525	11570	+ 0
50	Intr	11688	11819	+ 0
	Term	11918	12161	+ 0
	>4589443	/40267		
55	len =	1345	nex =	5
	Term	32380	32218	- 0
	Intr	32541	32485	- 0
	Intr	32744	32640	- 0
60	Intr	32964	32857	- 0

1491

	Init	33319	33125	-	0
	>4589443 /144066				
5	len =	1510	nex =	3	
	Term	37967	37540	-	0
	Intr	38213	38133	-	0
	Init	39044	38777	-	0
10	>4589444 /40174				
	len =	2138	nex =	7	
15	Init	5581	5696	+	0
	Intr	6020	6178	+	0
	Intr	6278	6453	+	0
	Intr	6565	6842	+	0
	Intr	6928	7063	+	0
20	Intr	7152	7263	+	0
	Term	7457	7718	+	0
	>4589444 /120707				
25	len =	1653	nex =	5	
	Init	68700	68826	+	0
	Intr	68906	69190	+	0
	Intr	69489	69569	+	0
30	Intr	69673	69767	+	0
	Term	69988	70352	+	0
	>4589444 /10687				
35	len =	582	nex =	2	
	Init	7159	7263	+	0
	Term	7457	7740	+	0
40	>4589444 /13338				
	len =	1435	nex =	2	
	Term	71110	70333	-	0
45	Init	71767	71187	-	0
	>4589444 /12533				
	len =	1612	nex =	3	
50	Term	72674	72238	-	0
	Intr	73073	72743	-	0
	Init	73849	73302	-	0
55	>4589444 /327				
	len =	1816	nex =	5	
	Init	74986	75176	+	0
60	Intr	75871	75973	+	0

2025 RELEASE UNDER E.O. 14176

1492

	Intr	76059	76148	+	0
	Intr	76241	76339	+	0
	Term	76490	76801	+	0
5	>4589445 /17344				
	len =	1030	nex =	2	
10	Init	1	139	+	0
	Term	623	1027	+	0
	>4589445 /38265				
15	len =	2260	nex =	9	
	Init	19710	19895	+	0
	Intr	20156	20228	+	0
	Intr	20310	20392	+	0
	Intr	20491	20593	+	0
20	Intr	20692	20848	+	0
	Intr	20926	21054	+	0
	Intr	21150	21236	+	0
	Intr	21401	21445	+	0
	Term	21645	21969	+	0
25	>4589445 /40038				
	len =	1410	nex =	1	
30	Sngl	23376	22781	-	0
	>4589445 /40717				
35	len =	1171	nex =	6	
	Term	59227	59185	-	0
	Intr	59432	59317	-	0
	Intr	59660	59559	-	0
	Intr	59846	59745	-	0
40	Intr	60036	59926	-	0
	Init	60355	60179	-	0
	>4589445 /6617				
45	len =	1852	nex =	4	
	Init	79257	79346	+	0
	Intr	79428	79703	+	0
	Intr	80505	80663	+	0
50	Term	80741	81108	+	0
	>4589446 /20765				
55	len =	2682	nex =	8	
	Term	14307	14077	-	0
	Intr	14445	14401	-	0
	Intr	14633	14514	-	0
	Intr	14989	14901	-	0
60	Intr	15212	15134	-	0

2025 RELEASE UNDER E.O. 14176

1493

	Intr	15616	15454	-	0
	Intr	15937	15858	-	0
	Init	16118	16011	-	0
5	>4589446 /38867				
	len =	1907	nex =	4	
10	Init	17211	17769	+	0
	Intr	18004	18141	+	0
	Intr	18241	18490	+	0
	Term	18625	19117	+	0
15	>4589450 /124077				
	len =	1690	nex =	6	
20	Init	6092	6223	+	0
	Intr	6597	6755	+	0
	Intr	7007	7256	+	0
	Intr	7347	7436	+	0
	Intr	7529	7617	+	0
	Term	7744	7770	+	0
25	>4589950 /6704				
	len =	2072	nex =	5	
30	Init	15942	16097	+	0
	Intr	16230	16315	+	0
	Intr	16673	16798	+	0
	Intr	16932	17059	+	0
	Term	17165	17489	+	0
35	>4589950 /152803				
	len =	611	nex =	1	
40	Sngl	23934	24544	+	0
	>4589950 /38094				
	len =	326	nex =	1	
45	Sngl	24221	24542	+	0
	>4589950 /20770				
50	len =	1643	nex =	4	
	Init	27913	28123	+	0
	Intr	28278	28494	+	0
	Intr	28594	28778	+	0
	Term	29119	29555	+	0
55	>4589950 /17664				
	len =	2673	nex =	4	
60	Init	30408	30899	+	0

2025 RELEASE UNDER E.O. 14176

1494

Intr	31702	32034	+	0
Intr	32610	32726	+	0
Term	32854	33080	+	0

5 >4589950 /21841

len = 1630 nex = 5

10	Init	43014	43406	+	0
	Intr	43536	43676	+	0
	Intr	43768	43908	+	0
	Intr	44011	44117	+	0
	Term	44298	44634	+	0

15 >4589950 /11283

len = 1270 nex = 2

20	Term	3662	3371	-	0
	Init	4626	4390	-	0

>4589950 /100317

25 len = 1255 nex = 2

	Term	3662	3416	-	0
	Init	4670	4390	-	0

30 >4589950 /41633

len = 569 nex = 1

	Sngl	50817	51385	+	0
--	------	-------	-------	---	---

35 >4589969 /120267

len = 610 nex = 2

40	Init	40423	40571	+	0
	Term	40797	41025	+	0

>4646215 /34289

45 len = 776 nex = 3

	Init	19221	19325	+	0
	Intr	19412	19533	+	0
	Term	19667	19993	+	0

50 >4646215 /9248

len = 867 nex = 1

55 Sngl 21032 20166 - 0

>4646229 /103197

len = 751 nex = 2

60 Init 3186 3298 + 0

Reference No. 2750-942P

					1495	0
	Term	3509	3709	+		
	>4646229	/19631				
5	len =	691	nex =	1		
	Sngl	56338	57028	+		0
	>4646229	/2267				
10	len =	2371	nex =	7		
	Init	59713	59825	+		0
	Intr	59944	59982	+		0
15	Intr	60076	60125	+		0
	Intr	60872	60983	+		0
	Intr	61359	61406	+		0
	Intr	61583	61713	+		0
20	Term	61829	61997	+		0
	>4646229	/6261				
	len =	408	nex =	1		
25	Sngl	64632	65028	+		0
	>4662609	/24161				
	len =	1541	nex =	5		
30	Init	110615	110883	+		0
	Intr	111111	111219	+		0
	Intr	111307	111437	+		0
	Intr	111538	111612	+		0
35	Term	111789	112155	+		0
	>4662609	/796				
	len =	618	nex =	2		
40	Init	110728	110883	+		0
	Term	111111	111219	+		0
	>4662609	/18583				
45	len =	1090	nex =	6		
	Init	11288	11484	+		0
	Intr	11708	11848	+		0
50	Intr	11922	12002	+		0
	Intr	12074	12198	+		0
	Intr	12253	12295	+		0
	Term	12327	12377	+		0
55	>4662609	/32558				
	len =	2255	nex =	5		
	Init	119155	119438	+		0
60	Intr	119536	119598	+		0

1496

Intr	120052	120472	+	0
Intr	120567	120718	+	0
Term	120832	121409	+	0

5 >4662609 /103288

len = 1570 nex = 6

10	Init	42796	42919	+	0
	Intr	43094	43245	+	0
	Intr	43358	43467	+	0
	Intr	43556	43641	+	0
	Intr	43725	43829	+	0
	Term	43910	44068	+	0

15 >4662609 /7803

len = 379 nex = 2

20	Term	58481	58268	-	0
	Init	58640	58572	-	0

>4662609 /34358

25	len =	873	nex =	1	
	Sngl	75547	76419	+	0

>4662628 /9376

30	len =	1657	nex =	4	
	Init	31339	31677	+	0
	Intr	31931	32188	+	0
35	Intr	32461	32664	+	0
	Term	32749	32995	+	0

>4662628 /154050

40	len =	1099	nex =	2	
	Term	36454	36398	-	0
	Init	37496	37156	-	0

45 >4662628 /20900

	len =	1390	nex =	1	
	Sngl	7708	9057	+	0

50 >4662637 /20182

len = 2859 nex = 7

55	Term	28149	27892	-	0
	Intr	28599	28528	-	0
	Intr	28829	28763	-	0
	Intr	29070	28932	-	0
	Intr	29862	29688	-	0
60	Intr	30330	30177	-	0

ORIGINAL DOCUMENT

1497

	Init	30750	30535	-	0
	>4662640	/103735			
5	len =	670	nex =	1	
	Sngl	1875	1670	-	0
	>4662640	/13193			
10	len =	741	nex =	1	
	Sngl	24672	23932	-	0
	>4662640	/108284			
15	len =	1342	nex =	2	
	Term	1875	1599	-	0
20	Init	2933	2230	-	0
	>4662640	/32647			
	len =	1474	nex =	2	
25	Term	1875	1671	-	0
	Init	3144	2230	-	0
	>4662640	/111177			
30	len =	370	nex =	1	
	Sngl	3147	2785	-	0
	>4662640	/32660			
35	len =	2432	nex =	3	
	Term	7037	6524	-	0
40	Intr	7476	7292	-	0
	Init	8955	8479	-	0
	>4678196	/1940			
45	len =	830	nex =	4	
	Term	30655	30442	-	0
	Intr	30869	30753	-	0
	Intr	31063	30949	-	0
50	Init	31271	31146	-	0
	>4678196	/30469			
	len =	858	nex =	4	
55	Term	30655	30458	-	0
	Intr	30869	30753	-	0
	Intr	31063	30949	-	0
	Init	31315	31146	-	0
60					

005120-9225560

1498

>4678196 /18195

```
len =      984    nex =      4
```

5	Term	30655	30384	-	0
	Intr	30869	30753	-	0
	Intr	31063	30949	-	0
	Init	31367	31146	-	0

10 >4678196 /994

```
len = 1046    nex = 4
```

	Term	7202	6888	-	0
15	Intr	7473	7304	-	0
	Intr	7664	7566	-	0
	Init	7933	7778	-	0

>4678219 /21113

```

20      len =      414      nex =      1

```

Snq1	16445	16858	+	0
------	-------	-------	---	---

25 >4678219 /124189

```
len = 1617    nex = 6
```

30	Init	17224	17328	+	0
	Intr	17546	17641	+	0
	Intr	17745	17806	+	0
	Intr	18172	18305	+	0
	Intr	18426	18510	+	0
	Term	18608	18840	+	0

35 >4678219 /10500

```
len =      970    nex =      3
```

40	Init	18935	19234	+	0
	Intr	19376	19429	+	0
	Term	19531	19898	+	0

>4678219 /41471

```
len = 1498    nex = 6
```

	Init	20212	20321	+	0
	Intr	20661	20780	+	0
50	Intr	20859	20904	+	0
	Intr	21091	21176	+	0
	Intr	21267	21396	+	0
	Term	21486	21709	+	0

55 >4678219 /1918

```
len = 1090    nex = 5
```

	Init	22504	22649	+	0
60	Intr	22728	22814	+	0

1499

Intr	22918	23020	+	0
Intr	23256	23322	+	0
Term	23414	23593	+	0

5 >4678258 /38927

len = 2446 nex = 8

Term	35527	35138	-	0
Intr	35817	35649	-	0
Intr	36025	35917	-	0
Intr	36381	36162	-	0
Intr	36617	36472	-	0
Intr	36786	36696	-	0
Intr	37244	37130	-	0
Init	37583	37343	-	0

>4678258 /2632

20 len = 2950 nex = 10

Term	38519	37849	-	0
Intr	38936	38848	-	0
Intr	39250	39065	-	0
Intr	39398	39336	-	0
Intr	39714	39532	-	0
Intr	39870	39822	-	0
Intr	40128	39990	-	0
Intr	40328	40247	-	0
Intr	40471	40416	-	0
Init	40791	40667	-	0

>4678258 /30342

35 len = 533 nex = 1

Sngl 88640 89172 + 0

>4678266 /19033

40 len = 577 nex = 1

Sngl 1664 2240 + 0

45 >4678266 /22956

len = 642 nex = 1

Sngl 4832 5473 + 0

50 >4678266 /13259

len = 2274 nex = 4

Init	73401	73742	+	0
Intr	73945	74637	+	0
Intr	74813	75124	+	0
Term	75360	75674	+	0

60 >4678266 /332

20190915 14:59:59

1500

len = 1090 nex = 1

5 Sngl 77266 76186 - 0

>4678266 /27474

len = 2792 nex = 10

10 Term 90126 89784 - 0

Intr 90502 90362 - 0

Intr 90835 90594 - 0

Intr 91028 90903 - 0

Intr 91211 91118 - 0

15 Intr 91432 91331 - 0

Intr 91850 91778 - 0

Intr 91990 91933 - 0

Intr 92372 92188 - 0

20 Init 92566 92480 - 0

>4678291 /19152

len = 1917 nex = 7

25 Init 17191 17291 + 0

Intr 17368 17833 + 0

Intr 17968 18085 + 0

Intr 18232 18341 + 0

Intr 18427 18512 + 0

30 Intr 18635 18780 + 0

Term 18898 19107 + 0

>4678291 /30751

35 len = 1344 nex = 7

Init 22561 22613 + 0

Intr 22732 22806 + 0

Intr 22896 22964 + 0

40 Intr 23038 23118 + 0

Intr 23250 23345 + 0

Intr 23435 23491 + 0

Term 23601 23896 + 0

45 >4678291 /33606

len = 446 nex = 1

Sngl 23601 23896 + 0

50 >4678291 /5493

len = 1270 nex = 2

55 Init 27193 27656 + 0

Term 27889 28459 + 0

>4678291 /37225

60 len = 2003 nex = 3

000150-0230000

1501

	Init	42970	43714	+	0
	Intr	43777	43917	+	0
	Term	44045	44522	+	0
5	>4678291 /21763				
	len =	933	nex =	3	
10	Init	46242	46546	+	0
	Intr	46860	46962	+	0
	Term	47050	47174	+	0
	>4678291 /25149				
15	len =	2552	nex =	8	
	Term	66537	66278	-	0
	Intr	66874	66788	-	0
20	Intr	67068	66979	-	0
	Intr	67487	67382	-	0
	Intr	67678	67584	-	0
	Intr	67904	67821	-	0
	Intr	68418	68343	-	0
25	Init	68829	68689	-	0
	>4678291 /15212				
	len =	2530	nex =	7	
30	Term	66537	66335	-	0
	Intr	66874	66788	-	0
	Intr	67068	66979	-	0
	Intr	67487	67382	-	0
35	Intr	67678	67584	-	0
	Intr	67904	67821	-	0
	Init	68422	68343	-	0
	>4678315 /34579				
40	len =	1990	nex =	5	
	Init	13656	13884	+	0
	Intr	13988	14067	+	0
45	Intr	14162	14239	+	0
	Intr	14367	14437	+	0
	Term	15188	15636	+	0
	>4678315 /787				
50	len =	324	nex =	1	
	Sngl	18778	19101	+	0
55	>4678315 /147241				
	len =	341	nex =	1	
60	Sngl	2959	3299	+	0

2025 RELEASE UNDER E.O. 14176

1502

>4678315 /34551

len = 1545 nex = 7

5	Term	41502	41430	-	0
	Intr	41669	41604	-	0
	Intr	41835	41771	-	0
	Intr	41994	41931	-	0
	Intr	42197	42088	-	0
10	Intr	42483	42396	-	0
	Init	42974	42809	-	0

>4678315 /9581

15 len = 1336 nex = 2

	Init	46422	46595	+	0
	Term	47332	47757	+	0

20 >4678340 /12236

len = 1954 nex = 5

	Term	22952	22722	-	0
25	Intr	23125	23045	-	0
	Intr	23705	23209	-	0
	Intr	24098	23940	-	0
	Init	24675	24184	-	0

30 >4678340 /26537

len = 2021 nex = 8

	Term	22952	22713	-	0
35	Intr	23125	23045	-	0
	Intr	23255	23209	-	0
	Intr	23445	23350	-	0
	Intr	23705	23639	-	0
	Intr	24098	23940	-	0
40	Intr	24254	24184	-	0
	Init	24733	24323	-	0

>4678340 /13870

45 len = 1640 nex = 3

	Term	3619	3273	-	0
	Intr	3881	3700	-	0
	Init	4912	4243	-	0

50 >4678371 /21034

len = 971 nex = 2

55	Init	109206	109301	+	0
	Term	109446	109715	+	0

>4678371 /113114

60 len = 892 nex = 2

2025 RELEASE UNDER E.O. 14176

1503

	Init	109206	109301	+	0
	Term	109446	109693	+	0
5	>4678371	/98073			
	len =	1111	nex =	3	
10	Term	41665	41333	-	0
	Intr	42181	42112	-	0
	Init	42443	42246	-	0
	>4678371	/41730			
15	len =	440	nex =	1	
	Sngl	48928	49367	+	0
	>4678705	/41543			
20	len =	1690	nex =	3	
	Init	11432	11814	+	0
	Intr	12077	12267	+	0
25	Term	12387	12550	+	0
	>4678705	/33399			
	len =	1690	nex =	7	
30	Init	14667	14750	+	0
	Intr	14858	15075	+	0
	Intr	15261	15331	+	0
	Intr	15428	15529	+	0
35	Intr	15646	15763	+	0
	Intr	15929	16064	+	0
	Term	16150	16346	+	0
	>4678705	/125324			
40	len =	2171	nex =	7	
	Init	16715	16857	+	0
	Intr	17191	17237	+	0
45	Intr	17759	17858	+	0
	Intr	18087	18150	+	0
	Intr	18237	18353	+	0
	Intr	18453	18561	+	0
	Term	18819	18885	+	0
50	>4678705	/24780			
	len =	791	nex =	2	
55	Init	25236	25351	+	0
	Term	25594	26026	+	0
	>4678705	/37881			
60	len =	2675	nex =	5	

1504

	Term	46430	45826	-	0
	Intr	46602	46534	-	0
	Intr	46744	46682	-	0
5	Intr	46896	46840	-	0
	Init	48500	47963	-	0
>4678705		/27536			
10	len =	281	nex =	1	
	Sngl	50278	49998	-	0
>4678705		/8207			
15	len =	1465	nex =	1	
	Sngl	51349	49885	-	0
>4678705		/15484			
20	len =	646	nex =	3	
	Term	68931	68742	-	0
25	Intr	69161	69058	-	0
	Init	69387	69316	-	0
>4680765		/19054			
30	len =	658	nex =	1	
	Sngl	103945	104602	+	0
>4680765		/12712			
35	len =	1157	nex =	4	
	Init	2888	2958	+	0
	Intr	3052	3180	+	0
40	Intr	3402	3584	+	0
	Term	3670	4024	+	0
>4680765		/29972			
45	len =	1551	nex =	6	
	Term	4691	4427	-	0
	Intr	4849	4793	-	0
	Intr	4976	4932	-	0
50	Intr	5238	5176	-	0
	Intr	5385	5322	-	0
	Init	5977	5762	-	0
>4680765		/112559			
55	len =	1614	nex =	5	
	Term	4691	4427	-	0
	Intr	4849	4793	-	0
60	Intr	5238	4932	-	0

1505

Intr	5385	5322	-	0
Init	6040	5762	-	0

>4680765 /33173

5	len =	2179	nex =	5
---	-------	------	-------	---

	Term	87951	87650	-	0
	Intr	88636	88413	-	0
10	Intr	88934	88817	-	0
	Intr	89343	89032	-	0
	Init	89828	89505	-	0

>4680765 /23348

15	len =	2238	nex =	5
----	-------	------	-------	---

	Term	87951	87633	-	0
	Intr	88636	88413	-	0
20	Intr	88934	88817	-	0
	Intr	89343	89032	-	0
	Init	89870	89505	-	0

>4680765 /7057

25	len =	3012	nex =	5
----	-------	------	-------	---

	Term	87951	87675	-	0
	Intr	88636	88413	-	0
30	Intr	88934	88817	-	0
	Intr	89343	89032	-	0
	Init	89750	89505	-	0

>4689466 /31988

35	len =	610	nex =	1	
	Sngl	44480	45086	+	0

>4691223 /19796

40	len =	1654	nex =	5
----	-------	------	-------	---

	Term	118925	118453	-	0
45	Intr	119112	119052	-	0
	Intr	119318	119145	-	0
	Intr	119652	119563	-	0
	Init	120106	119970	-	0

>4691223 /33058

50	len =	1390	nex =	2
----	-------	------	-------	---

	Init	120319	120641	+	0
55	Term	121277	121707	+	0

>4691223 /7104

60	len =	2650	nex =	7
----	-------	------	-------	---

1506

	Term	27184	26876	-	0
	Intr	27622	27443	-	0
	Intr	27988	27700	-	0
	Intr	28141	28074	-	0
5	Intr	28414	28229	-	0
	Intr	28667	28503	-	0
	Init	29523	28767	-	0
	>4691223		/97866		
10	len =	654	nex =	3	
	Term	53529	53241	-	0
	Intr	53668	53616	-	0
15	Init	53894	53745	-	0
	>4691223		/118329		
	len =	507	nex =	1	
20	Sngl	61751	62257	+	0
	>4691223		/29714		
25	len =	473	nex =	3	
	Term	67718	67641	-	0
	Intr	67897	67808	-	0
	Init	68113	67979	-	0
30	>4691223		/115554		
	len =	379	nex =	1	
35	Sngl	72403	72781	+	0
	>4691223		/20915		
	len =	3232	nex =	8	
40	Init	78459	78663	+	0
	Intr	79140	79266	+	0
	Intr	79400	79506	+	0
	Intr	79713	79874	+	0
45	Intr	79963	80109	+	0
	Intr	80186	80380	+	0
	Intr	80488	80666	+	0
	Term	80745	80975	+	0
50	>4691223		/9841		
	len =	523	nex =	1	
	Sngl	81514	80992	-	0
55	>4699904		/21639		
	len =	1521	nex =	2	
60	Term	29454	28950	-	0

1507

	Init	30470	29953	-	0
	>4699904	/30530			
5	len =	2040	nex =	4	
	Term	32070	31765	-	0
	Intr	32214	32162	-	0
	Intr	33121	33047	-	0
10	Init	33255	33215	-	0
	>4699904	/37701			
	len =	790	nex =	1	
15	Sngl	38588	37803	-	0
	>4699904	/30113			
20	len =	790	nex =	1	
	Sngl	38588	37804	-	0
	>4699904	/105950			
25	len =	412	nex =	1	
	Sngl	38588	38177	-	0
30	>4699904	/9278			
	len =	310	nex =	1	
	Sngl	38588	38288	-	0
35	>4699904	/8446			
	len =	1770	nex =	4	
40	Init	82618	82845	+	0
	Intr	83226	83301	+	0
	Intr	83389	83561	+	0
	Term	84124	84387	+	0
45	>4713943	/38549			
	len =	2170	nex =	6	
	Init	11662	11887	+	0
50	Intr	12188	12265	+	0
	Intr	12354	12614	+	0
	Intr	12703	13020	+	0
	Intr	13129	13287	+	0
	Term	13447	13623	+	0
55	>4713943	/19103			
	len =	3190	nex =	13	
60	Init	18821	18995	+	0

1508

	Intr	19606	19673	+	0
	Intr	19753	19791	+	0
	Intr	19928	20028	+	0
	Intr	20168	20280	+	0
5	Intr	20369	20468	+	0
	Intr	20547	20693	+	0
	Intr	20774	20834	+	0
	Intr	20925	21022	+	0
	Intr	21110	21252	+	0
10	Intr	21361	21444	+	0
	Intr	21529	21618	+	0
	Term	21702	22007	+	0
	>4713943		/10072		
15	len =	1233	nex =	6	
	Init	20810	20834	+	0
	Intr	20925	21022	+	0
20	Intr	21110	21252	+	0
	Intr	21361	21444	+	0
	Intr	21529	21618	+	0
	Term	21702	22042	+	0
25	>4713943		/21947		
	len =	1195	nex =	2	
	Init	28706	29150	+	0
30	Term	29228	29900	+	0
	>4713943		/16095		
	len =	706	nex =	1	
35	Sngl	69411	70116	+	0
	>4713943		/155207		
40	len =	1067	nex =	2	
	Init	73695	73912	+	0
	Term	74344	74761	+	0
45	>4725940		/20383		
	len =	687	nex =	1	
	Sngl	106710	107396	+	0
50	>4725940		/6554		
	len =	614	nex =	1	
55	Sngl	110711	111324	+	0
	>4725940		/92350		
60	len =	1090	nex =	4	

1509

	Term	17635	17313	-	0
	Intr	17915	17751	-	0
	Intr	18084	18006	-	0
	Init	18376	18165	-	0
5	>4725940 /28572				
	len =	537	nex =	1	
10	Sngl	96158	95622	-	0
	>4732167 /2296				
	len =	1813	nex =	5	
15	Init	45294	45510	+	0
	Intr	45586	45739	+	0
	Intr	46139	46198	+	0
	Intr	46531	46591	+	0
20	Term	46828	47106	+	0
	>4732168 /8387				
	len =	404	nex =	1	
25	Sngl	120482	120649	+	0
	>4732168 /38543				
30	len =	1971	nex =	6	
	Term	79492	79052	-	0
	Intr	79689	79588	-	0
	Intr	79895	79770	-	0
35	Intr	80126	79992	-	0
	Intr	80429	80313	-	0
	Init	81022	80821	-	0
	>4732168 /40690				
40	len =	670	nex =	2	
	Term	80429	80357	-	0
	Init	81025	80821	-	0
45	>4732169 /10692				
	len =	532	nex =	1	
50	Sngl	52688	52157	-	0
	>4732169 /25430				
	len =	550	nex =	1	
55	Sngl	52692	52146	-	0
	>4733952 /37506				
60	len =	2578	nex =	7	

1510

	Init	117152	117890	+	0
	Intr	117977	118056	+	0
	Intr	118424	118501	+	0
5	Intr	118902	118997	+	0
	Intr	119102	119227	+	0
	Intr	119304	119369	+	0
	Term	119460	119729	+	0
10	>4733952	/11975			
	len =	1396	nex =	2	
	Init	2301	2629	+	0
15	Term	3230	3696	+	0
	>4733952	/10295			
	len =	1244	nex =	3	
20	Init	4419	5051	+	0
	Intr	5122	5416	+	0
	Term	5495	5662	+	0
25	>4733952	/109289			
	len =	1232	nex =	3	
	Init	4431	5051	+	0
30	Intr	5122	5416	+	0
	Term	5495	5662	+	0
	>4733952	/36621			
35	len =	1918	nex =	6	
	Term	86725	86366	-	0
	Intr	87006	86815	-	0
	Intr	87271	87194	-	0
40	Intr	87467	87359	-	0
	Intr	87678	87564	-	0
	Init	88281	87885	-	0
	>4733952	/28602			
45	len =	599	nex =	1	
	Sngl	97194	97792	+	0
50	>4733952	/32795			
	len =	719	nex =	1	
	Sngl	97257	97975	+	0
55	>4733953	/106946			
	len =	473	nex =	1	
60	Sngl	118125	117810	-	0

1511

>4733953 /29385

5	len =	2099	nex =	6	
	Term	10729	10095	-	0
	Intr	10883	10802	-	0
	Intr	11125	11042	-	0
	Intr	11384	11217	-	0
10	Intr	11617	11461	-	0
	Init	12193	11723	-	0

>4733953 /15169

15	len =	1239	nex =	2	
	Term	16338	16088	-	0
	Init	17326	17040	-	0

>4733953 /2629

20	len =	1270	nex =	1	
	Sngl	16338	16093	-	0

>4733953 /39404

25	len =	379	nex =	1	
30	Sngl	17961	17583	-	0

>4733953 /19759

35	len =	2110	nex =	3	
	Init	20860	21383	+	0
	Intr	21459	21893	+	0
	Term	22639	22966	+	0

>4733953 /30753

40	len =	1171	nex =	4	
	Term	60229	59859	-	0
45	Intr	60602	60530	-	0
	Intr	60835	60709	-	0
	Init	61029	60968	-	0

>4733953 /37277

50	len =	2830	nex =	9	
	Init	61852	62588	+	0
	Intr	62673	62871	+	0
55	Intr	62958	63050	+	0
	Intr	63137	63265	+	0
	Intr	63340	63393	+	0
	Intr	63487	63564	+	0
	Intr	63686	63748	+	0
60	Intr	63865	63977	+	0

1512

	Term	64059	64359	+	0
	>4733953	/101924			
5	len =	730	nex =	3	
	Init	73	181	+	0
	Intr	262	357	+	0
	Term	434	801	+	0
10	>4733953	/32143			
	len =	1905	nex =	4	
15	Term	97964	97577	-	0
	Intr	98361	98053	-	0
	Intr	98605	98433	-	0
	Init	99481	99057	-	0
20	>4733957	/125929			
	len =	970	nex =	2	
	Term	7176	6924	-	0
25	Init	7349	7257	-	0
	>4733984	/26824			
	len =	2050	nex =	3	
30	Init	47232	47598	+	0
	Intr	47718	47928	+	0
	Term	48007	48272	+	0
35	>4733984	/29669			
	len =	1938	nex =	3	
	Init	47232	47598	+	0
40	Intr	47718	47928	+	0
	Term	48007	48229	+	0
	>4734003	/93895			
45	len =	533	nex =	1	
	Sngl	20795	21327	+	0
	>4734003	/32391			
50	len =	2026	nex =	6	
	Init	45023	45326	+	0
	Intr	45415	45959	+	0
55	Intr	46054	46140	+	0
	Intr	46221	46277	+	0
	Intr	46354	46798	+	0
	Term	46863	47048	+	0
60	>4734003	/9500			

1513

	len =	802	nex =	2	
	Term	70600	70084	-	0
5	Init	70885	70675	-	0
	>4734003 /28779				
	len =	1210	nex =	5	
10	Term	69956	69706	-	0
	Intr	70115	70043	-	0
	Intr	70462	70257	-	0
	Intr	70600	70537	-	0
15	Init	70912	70675	-	0
	>4753195 /5767				
	len =	895	nex =	2	
20	Term	102330	101848	-	0
	Init	102742	102437	-	0
	>4753195 /20865				
25	len =	1902	nex =	5	
	Term	24230	23691	-	0
	Intr	24471	24314	-	0
30	Intr	24713	24567	-	0
	Intr	25027	24796	-	0
	Init	25592	25383	-	0
	>4753195 /7837				
35	len =	2021	nex =	9	
	Init	41032	41313	+	0
	Intr	41510	41597	+	0
40	Intr	41693	41806	+	0
	Intr	41892	41974	+	0
	Intr	42064	42178	+	0
	Intr	42268	42359	+	0
	Intr	42443	42508	+	0
45	Intr	42608	42701	+	0
	Term	42827	43052	+	0
	>4753645 /42927				
50	len =	889	nex =	1	
	Sngl	15300	14412	-	0
	>4753645 /748				
55	len =	912	nex =	1	
	Sngl	15333	14422	-	0
60	>4753645 /11649				

2025 RELEASE UNDER E.O. 14176

1514

	len =	778	nex =	1	
5	Sngl	20284	19507	-	0
	>4753645	/18082			
	len =	1035	nex =	3	
10	Term	25637	25286	-	0
	Intr	26018	25864	-	0
	Init	26320	26110	-	0
	>4753645	/4311			
15	len =	991	nex =	4	
	Term	31713	31426	-	0
	Intr	31904	31800	-	0
20	Intr	32064	32003	-	0
	Init	32416	32195	-	0
	>4753645	/121013			
25	len =	1018	nex =	3	
	Term	33198	32799	-	0
	Intr	33586	33527	-	0
	Init	33816	33685	-	0
30	>4753645	/20847			
	len =	1904	nex =	8	
35	Term	41741	41372	-	0
	Intr	42028	41836	-	0
	Intr	42227	42105	-	0
	Intr	42440	42301	-	0
	Intr	42595	42527	-	0
40	Intr	42881	42686	-	0
	Intr	43051	42977	-	0
	Init	43275	43144	-	0
	>4755178	/7413			
45	len =	550	nex =	1	
	Sngl	23401	23949	+	0
50	>4755179	/37863			
	len =	2208	nex =	4	
	Term	38083	37560	-	0
55	Intr	38382	38183	-	0
	Intr	38861	38468	-	0
	Init	39767	39548	-	0
	>4755179	/33047			
60					

1515

	len =	477	nex =	1	
	Sngl	45523	45047	-	0
5	>4755179	/13002			
	len =	1676	nex =	2	
	Init	50115	50339	+	0
10	Term	50656	51087	+	0
	>4755179	/125387			
	len =	774	nex =	2	
15	Init	50164	50339	+	0
	Term	50656	50937	+	0
	>4755179	/40183			
20	len =	2251	nex =	7	
	Term	73491	73290	-	0
	Intr	73894	73775	-	0
25	Intr	74166	73981	-	0
	Intr	74402	74263	-	0
	Intr	75048	74800	-	0
	Intr	75246	75129	-	0
	Init	75540	75409	-	0
30	>4755179	/4933			
	len =	1523	nex =	0	
35	>4755179	/40305			
	len =	1342	nex =	1	
40	Sngl	82160	83501	+	0
	>4755179	/112223			
	len =	279	nex =	1	
45	Sngl	86044	85766	-	0
	>4755179	/230			
	len =	2782	nex =	8	
50	Term	90030	89591	-	0
	Intr	90265	90182	-	0
	Intr	90522	90366	-	0
	Intr	90939	90674	-	0
55	Intr	91320	91048	-	0
	Intr	91530	91419	-	0
	Intr	91829	91610	-	0
	Init	92372	91910	-	0
60	>4755179	/113514			

1516

	len =	2378	nex =	6	
	Term	7459	7240	-	0
5	Intr	7721	7576	-	0
	Intr	7889	7809	-	0
	Intr	8124	7978	-	0
	Intr	8629	8530	-	0
	Init	9601	9467	-	0
10	>4755185 /27930				
	len =	861	nex =	3	
15	Term	104138	103879	-	0
	Intr	104520	104451	-	0
	Init	104739	104586	-	0
	>4755185 /4047				
20	len =	885	nex =	1	
	Sngl	104763	103879	-	0
	>4755185 /5198				
25	len =	910	nex =	3	
	Term	104138	103878	-	0
	Intr	104520	104451	-	0
30	Init	104780	104586	-	0
	>4755185 /99800				
	len =	996	nex =	0	
35	>4755185 /33377				
	len =	1606	nex =	1	
40	Sngl	89482	90543	+	0
	>4757388 /34183				
	len =	996	nex =	3	
45	Term	686	30	-	0
	Intr	828	758	-	0
	Init	1025	911	-	0
50	>4757388 /20852				
	len =	1221	nex =	2	
	Init	26560	26673	+	0
55	Term	26774	27780	+	0
	>4757390 /37617				
60	len =	2548	nex =	3	

1517

Term	13877	13660	-	0
Intr	14019	13968	-	0
Init	16207	15422	-	0

5	>4757390	/109272		
	len =	528	nex =	1
	Sngl	37853	38380	+
10	>4757390	/100074		
	len =	519	nex =	1
15	Sngl	37862	38380	+
	>4757390	/103168		
	len =	1856	nex =	4
20	Init	41052	41145	+
	Intr	41501	41594	+
	Intr	42124	42386	+
	Term	42471	42907	+
25	>4757392	/2102		
	len =	3220	nex =	13
30	Term	18611	18194	-
	Intr	18807	18736	-
	Intr	19090	18977	-
	Intr	19281	19168	-
	Intr	19487	19387	-
35	Intr	19668	19590	-
	Intr	19823	19749	-
	Intr	19989	19927	-
	Intr	20445	20365	-
	Intr	20669	20544	-
40	Intr	21040	20916	-
	Intr	21231	21126	-
	Init	21413	21349	-
	>4757392	/31656		
45	len =	970	nex =	3
	Term	28052	27639	-
	Intr	28365	28148	-
50	Init	28605	28452	-
	>4757392	/96588		
	len =	590	nex =	2
55	Term	41513	41072	-
	Init	41654	41598	-
	>4757392	/13932		

60

20250612 09:16:00

1518

	len =	1822	nex =	5	
	Term	48918	48477	-	0
	Intr	49253	49006	-	0
5	Intr	49533	49327	-	0
	Intr	49777	49615	-	0
	Init	50298	50008	-	0
	>4757392 /36559				
10	len =	1821	nex =	4	
	Term	48918	48478	-	0
	Intr	49253	49006	-	0
15	Intr	49533	49327	-	0
	Init	49777	49615	-	0
	>4757392 /2204				
20	len =	550	nex =	1	
	Sngl	5387	4851	-	0
	>4757392 /9425				
25	len =	579	nex =	1	
	Sngl	57089	57667	+	0
30	>4757392 /111154				
	len =	1092	nex =	0	
	>4757392 /26899				
35	len =	2182	nex =	4	
	Init	73097	73405	+	0
	Intr	73958	74660	+	0
40	Intr	74754	74910	+	0
	Term	74995	75278	+	0
	>4757392 /12456				
45	len =	1995	nex =	4	
	Init	73287	73405	+	0
	Intr	73958	74660	+	0
	Intr	74754	74910	+	0
50	Term	74995	75281	+	0
	>4757392 /35237				
	len =	254	nex =	1	
55	Sngl	75025	75278	+	0
	>4757392 /16548				
60	len =	1845	nex =	5	

1519

	Term	85494	84859	-	0
	Intr	85695	85618	-	0
	Intr	86164	86108	-	0
5	Intr	86314	86260	-	0
	Init	86703	86408	-	0
	>4757395 /15932				
10	len =	801	nex =	2	
	Term	42705	42137	-	0
	Init	42937	42787	-	0
15	>4757395 /13267				
	len =	1911	nex =	4	
	Init	53364	54057	+	0
20	Intr	54543	54725	+	0
	Intr	54857	55024	+	0
	Term	55121	55274	+	0
	>4757395 /118150				
25	len =	1918	nex =	3	
	Init	55805	56142	+	0
	Intr	56228	56443	+	0
30	Term	57409	57722	+	0
	>4757396 /4410				
	len =	471	nex =	1	
35	Sngl	15964	16434	+	0
	>4757396 /115489				
40	len =	459	nex =	1	
	Sngl	15976	16434	+	0
	>4757399 /115876				
45	len =	341	nex =	1	
	Sngl	1954	2294	+	0
50	>4757400 /11763				
	len =	1275	nex =	2	
	Term	74376	74009	-	0
55	Init	75283	74843	-	0
	>4757400 /8330				
60	len =	2249	nex =	8	

1520

	Term	80687	80160	-	0
	Intr	80926	80866	-	0
	Intr	81202	81027	-	0
	Intr	81363	81306	-	0
5	Intr	81868	81764	-	0
	Intr	82022	81965	-	0
	Intr	82194	82108	-	0
	Init	82408	82279	-	0
10	>4757401 /924				
	len =	1438	nex =	3	
	Term	13076	12445	-	0
15	Intr	13565	13169	-	0
	Init	13882	13663	-	0
	>4757401 /103058				
20	len =	550	nex =	1	
	Sngl	22797	22249	-	0
	>4757401 /115850				
25	len =	614	nex =	1	
	Sngl	26499	25886	-	0
	>4757401 /27707				
	len =	716	nex =	2	
	Term	27638	27345	-	0
35	Init	28060	27979	-	0
	>4757401 /124616				
	len =	645	nex =	1	
40	Sngl	32214	31570	-	0
	>4757401 /27482				
45	len =	2410	nex =	6	
	Init	75050	75395	+	0
	Intr	75597	75744	+	0
	Intr	75850	76110	+	0
50	Intr	76313	76357	+	0
	Intr	76564	76732	+	0
	Term	76891	77454	+	0
	>4757403 /13058				
55	len =	1510	nex =	7	
	Term	15873	15797	-	0
	Intr	16028	15960	-	0
60	Intr	16200	16113	-	0

1521

	Intr	16374	16279	-	0
	Intr	16632	16519	-	0
	Intr	16870	16737	-	0
	Init	17303	16971	-	0
5	>4757403 /38239				
	len =	2056	nex =	7	
10	Term	15873	15307	-	0
	Intr	16028	15960	-	0
	Intr	16200	16113	-	0
	Intr	16374	16279	-	0
	Intr	16632	16519	-	0
15	Intr	16870	16737	-	0
	Init	17362	16971	-	0
	>4757403 /19537				
20	len =	1552	nex =	4	
	Term	19050	18584	-	0
	Intr	19389	19285	-	0
	Intr	19620	19513	-	0
25	Init	20135	19978	-	0
	>4757403 /24666				
	len =	1611	nex =	2	
30	Init	24264	25305	+	0
	Term	25397	25874	+	0
	>4757403 /39358				
35	len =	2877	nex =	8	
	Term	26486	25989	-	0
	Intr	26777	26718	-	0
40	Intr	26988	26867	-	0
	Intr	27249	27072	-	0
	Intr	27624	27359	-	0
	Intr	28030	27742	-	0
	Intr	28252	28115	-	0
45	Init	28406	28339	-	0
	>4757403 /39534				
	len =	554	nex =	1	
50	Sngl	32923	32370	-	0
	>4757403 /7337				
55	len =	503	nex =	1	
	Sngl	35283	34781	-	0
	>4757403 /40618				
60					

2025 RELEASE UNDER E.O. 14176

1522

	len =	1967	nex =	6	
	Init	51540	51679	+	0
	Intr	51794	51843	+	0
5	Intr	52329	52391	+	0
	Intr	52696	52786	+	0
	Intr	52894	52998	+	0
	Term	53300	53506	+	0
10	>4757403	/38488			
	len =	610	nex =	1	
15	Sngl	77818	78427	+	0
	>4757404	/103083			
	len =	143	nex =	1	
20	Sngl	32868	33010	+	0
	>4757405	/36106			
25	len =	2269	nex =	7	
	Term	8267	7900	-	0
	Intr	8590	8359	-	0
	Intr	8829	8685	-	0
	Intr	9156	8911	-	0
30	Intr	9449	9249	-	0
	Intr	9730	9637	-	0
	Init	10168	9818	-	0
	>4757405	/20648			
35	len =	518	nex =	1	
	Sngl	39529	39012	-	0
40	>4757405	/94317			
	len =	397	nex =	1	
	Sngl	41006	40610	-	0
45	>4757405	/100878			
	len =	1572	nex =	7	
50	Term	45130	44797	-	0
	Intr	45395	45330	-	0
	Intr	45558	45474	-	0
	Intr	45726	45633	-	0
	Intr	45884	45827	-	0
55	Intr	46200	46160	-	0
	Init	46368	46277	-	0
	>4757405	/5385			
60	len =	599	nex =	1	

1523

	Sngl	47192	46605	-	0
5	>4757405	/40773			
	len =	598	nex =	1	
	Sngl	47192	46606	-	0
10	>4757405	/30812			
	len =	1544	nex =	6	
15	Term	52068	51817	-	0
	Intr	52294	52205	-	0
	Intr	52456	52409	-	0
	Intr	52724	52555	-	0
	Intr	53078	53000	-	0
	Init	53360	53196	-	0
20	>4757405	/16619			
	len =	761	nex =	2	
25	Init	56271	56389	+	0
	Term	56759	57031	+	0
	>4757405	/3049			
30	len =	2274	nex =	5	
	Init	56278	57300	+	0
	Intr	57384	57484	+	0
	Intr	57586	57669	+	0
35	Intr	57761	57871	+	0
	Term	58094	58551	+	0
	>4757406	/19244			
40	len =	1177	nex =	2	
	Init	25061	25439	+	0
	Term	25638	26237	+	0
45	>4757406	/35710			
	len =	1589	nex =	8	
50	Init	31143	31329	+	0
	Intr	31449	31591	+	0
	Intr	31679	31786	+	0
	Intr	31888	31932	+	0
	Intr	32002	32067	+	0
	Intr	32152	32212	+	0
55	Intr	32296	32342	+	0
	Term	32424	32731	+	0
	>4757406	/97340			
60	len =	2530	nex =	5	

1524

	Init	33489	33681	+	0
	Intr	34354	34432	+	0
	Intr	34531	34598	+	0
5	Intr	34773	34839	+	0
	Term	35246	35497	+	0
	>4757407		/145523		
10	len =	522	nex =	2	
	Init	14432	14663	+	0
	Term	14777	14953	+	0
15	>4757407		/16944		
	len =	1228	nex =	5	
	Term	35221	34886	-	0
20	Intr	35427	35369	-	0
	Intr	35569	35507	-	0
	Intr	35739	35676	-	0
	Init	36113	35815	-	0
25	>4757407		/18624		
	len =	562	nex =	1	
	Sngl	42900	43461	+	0
30	>4757407		/19127		
	len =	2025	nex =	5	
35	Term	44665	44410	-	0
	Intr	44838	44752	-	0
	Intr	45182	44922	-	0
	Intr	45743	45338	-	0
	Init	46434	45860	-	0
40	>4757407		/41161		
	len =	1106	nex =	3	
45	Init	5273	5534	+	0
	Intr	5627	5929	+	0
	Term	6019	6378	+	0
	>4757407		/11077		
50	len =	466	nex =	1	
	Sngl	6030	6495	+	0
55	>4757407		/111727		
	len =	3141	nex =	5	
	Init	75926	76044	+	0
60	Intr	76156	76326	+	0

2025 RELEASE UNDER E.O. 14176

1525

Intr	76407	76463	+	0
Intr	76541	76605	+	0
Term	76716	76840	+	0

5 >4757407 /12484

```
len = 1893    nex = 4
```

10	Term	77674	77314	-	0
	Intr	77970	77850	-	0
	Intr	78287	78047	-	0
	Init	79206	78928	-	0

>4757407 /26596

```

15      len =      1953      nex =          5

```

Term	82383	82097	-	0
Intr	82758	82519	-	0
Intr	83472	83378	-	0
Intr	83776	83689	-	0
Init	84049	84002	-	0

>4757409 /6166

```
len = 2028    nex = 3
```

Init	2893	3040	+	0
Intr	3127	3389	+	0
Term	3462	4256	+	0

>4757409 /17962

```
len = 790    nex = 2
```

Term	4534	4275	-	0
Init	5064	4726	-	0

>4757410 /13767

```

40      len =      2170      nex =          5

```

Init	11323	11581	+	0
Intr	11799	11940	+	0
Intr	12035	12088	+	0
Intr	12453	12544	+	0
Term	12656	12882	+	0

>4757410 /104278

```
50      len =      970      nex =      1
```

Snq1	14335	13368	-	0
------	-------	-------	---	---

55 >4757410 /5180

```
len =      626    nex =      2
```

	Init	37202	37397	+	0
60	Term	37493	37827	+	0

1526

>4757410 /92670

```

5      len =      1849      nex =          2
      Term  42000  41613      -          0
      Init  43461  43210      -          0

```

>4757410 /29464

```
len =      895    nex =          1
Snq1 50057    50951      +          0
```

15 >4757410 /3819

```
len =      2093      nex =          2

Init  59901      60514      +          0
Term  60601      60838      +          0
```

>4757410 /34479

```

25      len =      1298      nex =          3
        Init 59369 59822      +      0
        Intr 59901 60514      +      0
        Term 60601 60666      +      0

```

30 >4757410 /23740

```
len =      1303      nex =          3
Init    59614    59822      +          0
Intr    59901    60514      +          0
Term    60601    60916      +          0
```

>4757410 /36248

```

40      len =      616      nex =      2
      Init 60322  60514      +      0
      Term 60601  60918      +      0

```

45 >4757410 /101693

```
len =      2146      nex =      6

Init    67945    68152    +    0
Intr    68845    68971    +    0
Intr    69127    69273    +    0
Intr    69364    69444    +    0
Intr    69538    69683    +    0
Term    69773    70080    +    0
```

>4757410 /123200

```

len =      284    nex =      1
60      Sngl  73051  72768      -      0

```

1527

>4757411 /99062

```

5      len =      348      nex =      1
      Snq1 14302 14649      +      0

```

>4757411 /38286

10	len =	2413	nex =	5	
	Init	35325	35620	+	0
	Intr	36178	36431	+	0
	Intr	36689	36824	+	0
15	Intr	37137	37198	+	0
	Term	37290	37737	+	0

>4757411 /143077

```
20      len =      774      nex =      3
      Term 48735 48361 - 0
      Intr 48922 48839 - 0
      Init 49134 48982 - 0
```

>4757413 /39613

	len =	1930	nex =	6	
30	Term	10486	10044	-	0
	Intr	10673	10581	-	0
	Intr	10937	10794	-	0
	Intr	11196	11110	-	0
	Intr	11493	11268	-	0
35	Init	11973	11691	-	0

>4757413 /37104

```

40      len =    1352    nex =        2
        Init 37253    37571      +      0
        Term 37707    38604      +      0

```

>4757413 /37513

len =	1375	nex =	2		
Init	37253	37571	+	0	
Term	37707	38627	+	0	

>4757413 /9683

```

len =      1290    nex =      1
55      Sngl 40182  39244      -      0

```

>4757414 /94738

```

60      len =      731      nex =      1

```


					1528
	Sngl	15443	14713	-	0
	>4757414	/1185			
5	len =	1812	nex =	2	
	Init	21364	21435	+	0
	Term	21537	22038	+	0
10	>4757414	/113827			
	len =	443	nex =	1	
15	Sngl	35589	36031	+	0
	>4757414	/117748			
	len =	615	nex =	2	
20	Init	41281	41606	+	0
	Term	41845	41895	+	0
	>4757414	/6463			
25	len =	1403	nex =	3	
	Init	41281	41606	+	0
	Intr	41845	41895	+	0
	Term	42583	42683	+	0
30	>4757414	/23733			
	len =	1161	nex =	1	
35	Sngl	43908	42748	-	0
	>4757414	/32274			
	len =	1690	nex =	5	
40	Init	44451	44575	+	0
	Intr	44666	44762	+	0
	Intr	44860	44969	+	0
	Intr	45357	45431	+	0
45	Term	45541	45739	+	0
	>4757414	/17361			
	len =	167	nex =	1	
50	Sngl	67827	67661	-	0
	>4757414	/21721			
55	len =	4353	nex =	15	
	Init	79949	80134	+	0
	Intr	80262	80364	+	0
	Intr	80481	80545	+	0
60	Intr	80685	80760	+	0

2025 RELEASE UNDER E.O. 14176

1529

	Intr	81080	81144	+	0
	Intr	81467	81590	+	0
	Intr	81701	81826	+	0
	Intr	81943	81993	+	0
5	Intr	82107	82175	+	0
	Intr	82394	82502	+	0
	Intr	82646	82725	+	0
	Intr	82828	82910	+	0
	Intr	82992	83037	+	0
10	Intr	83119	83293	+	0
	Term	83381	83643	+	0
>4757415		/43010			
15	len =	2179	nex =	5	
	Init	110	480	+	0
	Intr	969	1054	+	0
	Intr	1131	1436	+	0
20	Intr	1538	1621	+	0
	Term	1706	2288	+	0
>4757415		/33108			
25	len =	1990	nex =	5	
	Term	18813	18525	-	0
	Intr	19074	18908	-	0
	Intr	19639	19580	-	0
30	Intr	19870	19724	-	0
	Init	20507	20200	-	0
>4757415		/35741			
35	len =	2319	nex =	7	
	Term	21262	20642	-	0
	Intr	21516	21346	-	0
	Intr	21769	21603	-	0
40	Intr	21961	21868	-	0
	Intr	22248	22040	-	0
	Intr	22485	22334	-	0
	Init	22960	22702	-	0
45	>4757415	/94723			
	len =	910	nex =	3	
	Term	8317	8042	-	0
50	Intr	8742	8423	-	0
	Init	8946	8856	-	0
>4757417		/12993			
55	len =	2378	nex =	8	
	Init	57794	57865	+	0
	Intr	57948	58085	+	0
	Intr	58174	58272	+	0
60	Intr	58373	58445	+	0

2025 RELEASE UNDER E.O. 14176

1530

Intr	58911	58999	+	0
Intr	59128	59211	+	0
Intr	59310	59369	+	0
Term	59441	59775	+	0

5

>4757417 /40272

```
len = 2254    nex = 8
```

10	Init	57533	57711	+	0
	Intr	57794	57865	+	0
	Intr	57948	58085	+	0
	Intr	58174	58272	+	0
	Intr	58373	58445	+	0
15	Intr	59128	59211	+	0
	Intr	59310	59369	+	0
	Term	59441	59786	+	0

>4757660 /34549

```
20      len =      2683      nex =      8
```

	Init	17650	17772	+	0
25	Intr	18075	18155	+	0
	Intr	18244	18332	+	0
	Intr	18448	18522	+	0
	Intr	18662	18726	+	0
	Intr	18812	18883	+	0
	Intr	18993	19430	+	0
30	Term	19524	20051	+	0

>4757660 /34984

```
len = 1570    nex = 6
```

	Init	23805	24076	+	0
	Intr	24164	24212	+	0
	Intr	24293	24419	+	0
	Intr	24566	24636	+	0
40	Intr	24733	24805	+	0
	Term	25086	25372	+	0

>4757660 /18595

```
45      len =      757      nex =      1
          Snq1 3045      3801      +      0
```

>4757660 /6179

```
50      len =      1773      nex =      5
      Term 38593 38287 - 0
      Intr 39063 38802 - 0
55      Intr 39389 39159 - 0
      Intr 39630 39497 - 0
      Init 40059 39789 - 0
```

>4757660 /31508

60

1531

	len =	1017	nex =	2	
	Term	41278	40970	-	0
5	Init	41687	41426	-	0
	>4757660 /295				
	len =	1534	nex =	5	
10	Term	41278	41051	-	0
	Intr	41687	41426	-	0
	Intr	42011	41778	-	0
	Intr	42219	42089	-	0
15	Init	42584	42322	-	0
	>4757660 /24456				
	len =	1006	nex =	2	
20	Term	45205	44865	-	0
	Init	45691	45430	-	0
	>4757660 /13876				
25	len =	1838	nex =	5	
	Term	45205	44861	-	0
	Intr	45691	45430	-	0
	Intr	46019	45783	-	0
30	Intr	46265	46135	-	0
	Init	46698	46390	-	0
	>4757660 /30015				
35	len =	1879	nex =	5	
	Term	47388	47089	-	0
	Intr	47741	47486	-	0
	Intr	48105	47875	-	0
40	Intr	48353	48223	-	0
	Init	48967	48677	-	0
	>4757660 /25120				
45	len =	1990	nex =	5	
	Term	47388	47089	-	0
	Intr	47741	47486	-	0
	Intr	48105	47875	-	0
50	Intr	48353	48223	-	0
	Init	49075	48677	-	0
	>4757660 /34896				
55	len =	1999	nex =	5	
	Term	54992	54648	-	0
	Intr	55353	55098	-	0
	Intr	55706	55473	-	0
60	Intr	55926	55796	-	0

2025-05-01 10:00:00

1532

Init	56646	56224	-	0
------	-------	-------	---	---

>4757660	/13762
----------	--------

5	len =	1939	nex =	7	
---	-------	------	-------	---	--

Term	69818	69378	-	0
------	-------	-------	---	---

Intr	69980	69915	-	0
------	-------	-------	---	---

Intr	70123	70059	-	0
------	-------	-------	---	---

10	Intr	70269	70206	-	0
----	------	-------	-------	---	---

Intr	70491	70382	-	0
------	-------	-------	---	---

Intr	70666	70579	-	0
------	-------	-------	---	---

Init	71316	71035	-	0
------	-------	-------	---	---

15	>4757660	/2663
----	----------	-------

len =	589	nex =	1
-------	-----	-------	---

Sngl	85423	84835	-	0
------	-------	-------	---	---

20	>4757660	/143795
----	----------	---------

len =	550	nex =	1
-------	-----	-------	---

25	Sngl	9114	8573	-	0
----	------	------	------	---	---

>4757661	/33846
----------	--------

len =	998	nex =	0
-------	-----	-------	---

30	>4757661	/18244
----	----------	--------

len =	1550	nex =	4
-------	------	-------	---

35	Init	20501	20937	+	0
----	------	-------	-------	---	---

Intr	21063	21205	+	0
------	-------	-------	---	---

Intr	21578	21697	+	0
------	-------	-------	---	---

Term	21792	22050	+	0
------	-------	-------	---	---

40	>4757661	/12689
----	----------	--------

len =	1820	nex =	3
-------	------	-------	---

Term	52653	52143	-	0
------	-------	-------	---	---

45	Intr	53031	52754	-	0
----	------	-------	-------	---	---

Init	53962	53118	-	0
------	-------	-------	---	---

>4757661	/37307
----------	--------

50	len =	1947	nex =	5
----	-------	------	-------	---

Init	79034	79330	+	0
------	-------	-------	---	---

Intr	79685	79897	+	0
------	-------	-------	---	---

Intr	79984	80132	+	0
------	-------	-------	---	---

55	Intr	80266	80542	+	0
----	------	-------	-------	---	---

Term	80634	80980	+	0
------	-------	-------	---	---

>4757661	/38874
----------	--------

60	len =	1550	nex =	3
----	-------	------	-------	---

1533

	Init	88004	88300	+	0
	Intr	88888	89103	+	0
	Term	89479	89553	+	0
5	>4757661 /29823				
	len =	1270	nex =	2	
10	Term	95428	95219	-	0
	Init	96487	96281	-	0
	>4757662 /151497				
15	len =	374	nex =	1	
	Sngl	105041	105406	+	0
	>4757662 /33355				
20	len =	764	nex =	1	
	Sngl	105041	105804	+	0
25	>4757662 /10525				
	len =	959	nex =	2	
	Term	107176	106447	-	0
30	Init	107405	107251	-	0
	>4757662 /40979				
	len =	2536	nex =	1	
35	Sngl	11058	12235	+	0
	>4757662 /3858				
40	len =	883	nex =	2	
	Term	127589	127180	-	0
	Init	128062	127787	-	0
45	>4757662 /25577				
	len =	641	nex =	1	
	Sngl	15911	16551	+	0
50	>4757662 /13756				
	len =	2333	nex =	6	
55	Init	26705	26948	+	0
	Intr	27713	27814	+	0
	Intr	27911	28038	+	0
	Intr	28342	28470	+	0
	Intr	28565	28632	+	0
60	Term	28747	29037	+	0

1534

>4757662 /19079

```
len =      1931      nex =          4

Term  35291  34651      -      0
Intr  35971  35763      -      0
Intr  36176  36076      -      0
Init  36581  36264      -      0
```

>4757662 /206217

```
len =      475    nex =      1
Sngl  51723    51806      +      0
```

>4757662 /41103

```
len =      2145      nex =      5

Term 55687 54683 - 0
Intr 55901 55794 - 0
Intr 56205 56003 - 0
Intr 56423 56280 - 0
Init 56827 56682 - 0
```

>4757662 /10861

```
len =      1789      nex =          5

Term 73379 73086 - 0
Intr 73578 73468 - 0
Intr 73781 73664 - 0
Intr 73905 73869 - 0
Init 74874 74703 - 0
```

>4757662 /12880

```
len =      4180      nex =      16

Term  75805  75445      -      0
Intr  76039  75900      -      0
Intr  76358  76118      -      0
Intr  76502  76452      -      0
Intr  76821  76596      -      0
Intr  77212  77142      -      0
Intr  77417  77298      -      0
Intr  77706  77512      -      0
Intr  77866  77802      -      0
Intr  78048  77958      -      0
Intr  78257  78196      -      0
Intr  78415  78365      -      0
Intr  78718  78617      -      0
Intr  78917  78810      -      0
Intr  79200  79137      -      0
Init  79624  79320      -      0
```

>4757662 /8333

```
len = 587    nex = 1
```


1536

Intr	16601	16514	-	0
Intr	16705	16680	-	0
Init	16918	16792	-	0

5 >4760247 /7361

len = 2251 nex = 6

10	Init	1746	2458	+	0
	Intr	2735	2837	+	0
	Intr	2918	3021	+	0
	Intr	3285	3376	+	0
	Intr	3462	3559	+	0
	Term	3647	3996	+	0

15 >4760247 /102245

len = 465 nex = 1

20 Sngl 1766 2230 + 0

>4760247 /16412

25 len = 2700 nex = 12

	Term	16228	15990	-	0
	Intr	16403	16313	-	0
	Intr	16601	16514	-	0
	Intr	17018	16792	-	0
30	Intr	17212	17097	-	0
	Intr	17537	17282	-	0
	Intr	17721	17634	-	0
	Intr	17870	17799	-	0
	Intr	18036	17928	-	0
35	Intr	18186	18128	-	0
	Intr	18460	18397	-	0
	Init	18689	18561	-	0

40 >4760247 /36952

len = 2304 nex = 5

	Term	6330	6048	-	0
	Intr	7503	7393	-	0
45	Intr	7682	7577	-	0
	Intr	7899	7771	-	0
	Init	8351	7975	-	0

50 >4760411 /33299

len = 2250 nex = 9

	Term	21949	21710	-	0
	Intr	22229	22137	-	0
55	Intr	22409	22356	-	0
	Intr	22653	22513	-	0
	Intr	22805	22740	-	0
	Intr	23090	22889	-	0
	Intr	23430	23376	-	0
60	Intr	23651	23590	-	0

00150-9425000

1537

	Init	23959	23734	-	0
	>4760411	/16491			
5	len =	1975	nex =	4	
	Init	4186	4655	+	0
	Intr	4887	5035	+	0
10	Intr	5129	5239	+	0
	Term	5334	6160	+	0
	>4760411	/120348			
15	len =	653	nex =	1	
	Sngl	45700	45048	-	0
	>4760411	/31883			
20	len =	152	nex =	1	
	Sngl	6057	6208	+	0
	>4760411	/36997			
25	len =	2050	nex =	8	
	Term	67189	66970	-	0
30	Intr	67456	67326	-	0
	Intr	67716	67588	-	0
	Intr	67946	67800	-	0
	Intr	68253	68062	-	0
	Intr	68522	68412	-	0
35	Intr	68705	68610	-	0
	Init	69013	68843	-	0
	>4760411	/30022			
40	len =	3315	nex =	6	
	Term	73699	73294	-	0
	Intr	74021	73925	-	0
	Intr	74270	74140	-	0
45	Intr	74416	74358	-	0
	Intr	76009	75544	-	0
	Init	76608	76282	-	0
	>4760411	/38193			
50	len =	1474	nex =	1	
	Sngl	77039	78512	+	0
	>4761801	/93707			
55	len =	619	nex =	1	
	Sngl	143737	143119	-	0
60	>4761801	/115966			

1538

	len =	649	nex =	1	
5	Sngl	151005	151653	+	0
	>4761801	/10618			
	len =	1736	nex =	5	
10	Init	168932	169123	+	0
	Intr	169266	169370	+	0
	Intr	169478	169788	+	0
	Intr	169866	170051	+	0
	Term	170355	170667	+	0
15	>4761801	/30044			
	len =	2558	nex =	5	
20	Init	26588	27075	+	0
	Intr	27771	27904	+	0
	Intr	28035	28462	+	0
	Intr	28545	28676	+	0
	Term	28863	29145	+	0
25	>4761801	/103581			
	len =	924	nex =	1	
30	Sngl	40802	39879	-	0
	>4761801	/96744			
	len =	1859	nex =	6	
35	Init	46090	46263	+	0
	Intr	46495	46564	+	0
	Intr	46666	46725	+	0
	Intr	46946	47060	+	0
40	Intr	47328	47443	+	0
	Term	47528	47948	+	0
	>4761801	/27650			
45	len =	643	nex =	2	
	Init	47359	47443	+	0
	Term	47528	47991	+	0
50	>4761801	/34620			
	len =	4450	nex =	14	
	Term	83154	82922	-	0
55	Intr	83294	83234	-	0
	Intr	83538	83451	-	0
	Intr	83672	83607	-	0
	Intr	84088	83988	-	0
	Intr	84594	84487	-	0
60	Intr	85379	85289	-	0

2025 RELEASE UNDER E.O. 14176

1539

	Intr	85690	85604	-	0
	Intr	85964	85841	-	0
	Intr	86148	86057	-	0
	Intr	86290	86229	-	0
5	Intr	86558	86461	-	0
	Intr	87054	86904	-	0
	Init	87364	87194	-	0
	>4775266 /119956				
10	len =	1411	nex =	2	
	Init	17383	17754	+	0
	Term	18006	18793	+	0
15	>4775266 /20805				
	len =	2816	nex =	9	
20	Init	19777	19920	+	0
	Intr	20186	20444	+	0
	Intr	20536	20742	+	0
	Intr	20819	21069	+	0
	Intr	21162	21486	+	0
25	Intr	21579	21842	+	0
	Intr	21927	22048	+	0
	Intr	22133	22252	+	0
	Term	22327	22592	+	0
30	>4775266 /20161				
	len =	557	nex =	1	
	Sngl	25109	24553	-	0
35	>4775266 /33023				
	len =	1956	nex =	4	
40	Term	27412	26849	-	0
	Intr	27825	27714	-	0
	Intr	28452	28152	-	0
	Init	28804	28549	-	0
45	>4775266 /10077				
	len =	715	nex =	2	
	Term	38914	38632	-	0
50	Init	39346	39105	-	0
	>4775266 /18040				
	len =	801	nex =	1	
55	Sngl	44171	44971	+	0
	>4775266 /7221				
60	len =	1630	nex =	6	

2025 RELEASE UNDER E.O. 14176

1540

	Term	4598	4139	-	0
	Intr	4777	4686	-	0
	Intr	4950	4865	-	0
5	Intr	5215	5035	-	0
	Intr	5520	5300	-	0
	Init	5766	5592	-	0
	>4803878 /4897				
10	len =	574	nex =	1	
	Sngl	22484	21911	-	0
	>4803878 /39544				
15	len =	1063	nex =	3	
	Term	38379	37933	-	0
20	Intr	38854	38542	-	0
	Init	38995	38931	-	0
	>4803878 /37896				
25	len =	1436	nex =	3	
	Term	38379	37680	-	0
	Intr	38854	38542	-	0
	Init	39115	38931	-	0
30	>4803878 /14132				
	len =	1582	nex =	4	
35	Term	38379	37679	-	0
	Intr	38854	38542	-	0
	Intr	39082	38931	-	0
	Init	39260	39188	-	0
40	>4803878 /4936				
	len =	537	nex =	1	
	Sngl	45234	44698	-	0
45	>4803878 /2725				
	len =	1431	nex =	4	
50	Term	51078	50648	-	0
	Intr	51328	51226	-	0
	Intr	51702	51580	-	0
	Init	52078	51914	-	0
55	>4803878 /10780				
	len =	1615	nex =	3	
	Init	52264	52570	+	0
60	Intr	52981	53054	+	0

1541

	Term	53516	53878	+	0
	>4803878 /40875				
5	len =	817	nex =	1	
	Sngl	65721	66537	+	0
	>4803878 /117183				
10	len =	1607	nex =	4	
	Init	73114	73395	+	0
	Intr	73476	73685	+	0
15	Intr	74219	74295	+	0
	Term	74422	74720	+	0
	>4803878 /19202				
20	len =	1552	nex =	4	
	Term	81492	81413	-	0
	Intr	81857	81669	-	0
	Intr	82005	81947	-	0
25	Init	82964	82666	-	0
	>4803878 /29968				
	len =	1375	nex =	5	
30	Term	83510	83314	-	0
	Intr	83688	83599	-	0
	Intr	83874	83772	-	0
	Intr	84042	83962	-	0
35	Init	84361	84132	-	0
	>4803878 /21256				
	len =	2506	nex =	10	
40	Init	88965	89156	+	0
	Intr	89749	89846	+	0
	Intr	89927	89969	+	0
	Intr	90054	90115	+	0
45	Intr	90200	90256	+	0
	Intr	90345	90380	+	0
	Intr	90639	90711	+	0
	Intr	90777	90848	+	0
	Intr	90945	91048	+	0
50	Term	91150	91470	+	0
	>4803878 /30884				
	len =	1120	nex =	5	
55	Term	91871	91785	-	0
	Intr	92048	91974	-	0
	Intr	92193	92137	-	0
	Intr	92413	92291	-	0
60	Init	92898	92778	-	0

000100-9225500

1542

>4803878 /6211

5	len =	1416	nex =	5	
	Term	91871	91565	-	0
	Intr	92048	91974	-	0
	Intr	92193	92137	-	0
	Intr	92413	92291	-	0
10	Init	92980	92778	-	0

>4803878 /118823

15	len =	1433	nex =	5	
	Term	91871	91597	-	0
	Intr	92048	91974	-	0
	Intr	92193	92137	-	0
	Intr	92413	92291	-	0
20	Init	93029	92778	-	0

>4803909 /33591

25	len =	2548	nex =	7	
	Init	25320	25450	+	0
	Intr	25545	25674	+	0
	Intr	25969	26075	+	0
	Intr	26572	26706	+	0
30	Intr	26804	26907	+	0
	Intr	27030	27130	+	0
	Term	27226	27508	+	0

>4803909 /101918

35	len =	2004	nex =	7	
	Init	25331	25450	+	0
	Intr	25545	25674	+	0
40	Intr	25969	26075	+	0
	Intr	26572	26706	+	0
	Intr	26804	26907	+	0
	Intr	27030	27130	+	0
45	Term	27226	27334	+	0

>4803909 /35385

	len =	1648	nex =	2	
50	Init	34510	35420	+	0
	Term	35503	36157	+	0

>4803909 /36399

55	len =	1999	nex =	5	
	Term	74683	74273	-	0
	Intr	75005	74768	-	0
	Intr	75269	75203	-	0
60	Intr	75560	75398	-	0

1543

	Init	76271	75939	-	0
	>4803909	/14382			
5	len =	572	nex =	1	
	Sngl	98753	98182	-	0
	>4803919	/7141			
10	len =	470	nex =	1	
	Sngl	10681	11150	+	0
15	>4803919	/20090			
	len =	638	nex =	3	
20	Term	11473	11254	-	0
	Intr	11717	11664	-	0
	Init	11891	11797	-	0
	>4803919	/19643			
25	len =	515	nex =	1	
	Sngl	18105	18619	+	0
	>4803919	/9942			
30	len =	628	nex =	2	
	Term	18385	18182	-	0
	Init	18466	18413	-	0
35	>4803919	/42149			
	len =	503	nex =	1	
40	Sngl	18515	18013	-	0
	>4803919	/37778			
	len =	570	nex =	1	
45	Sngl	19660	20229	+	0
	>4803919	/39351			
50	len =	2416	nex =	9	
	Init	21856	22144	+	0
	Intr	22554	22580	+	0
	Intr	22672	22741	+	0
55	Intr	22863	22986	+	0
	Intr	23146	23230	+	0
	Intr	23321	23453	+	0
	Intr	23544	23717	+	0
	Intr	23785	23867	+	0
60	Term	23953	24271	+	0

00000000-00000000

1544

>4803919 /40766

5	len =	2062	nex =	8	
	Term	24902	24659	-	0
	Intr	25064	24993	-	0
	Intr	25345	25142	-	0
	Intr	25530	25453	-	0
10	Intr	25707	25619	-	0
	Intr	25854	25804	-	0
	Intr	26112	25945	-	0
	Init	26720	26568	-	0

15 >4803919 /19342

	len =	2750	nex =	11	
	Term	27185	26968	-	0
20	Intr	27343	27269	-	0
	Intr	27502	27437	-	0
	Intr	27642	27589	-	0
	Intr	27860	27729	-	0
	Intr	28170	28094	-	0
25	Intr	28344	28266	-	0
	Intr	28920	28805	-	0
	Intr	29072	29008	-	0
	Intr	29223	29158	-	0
	Init	29717	29489	-	0

>4803919 /35643

	len =	757	nex =	1	
35	Sngl	38824	39580	+	0

>4803919 /39721

	len =	2799	nex =	9	
40	Init	40435	40910	+	0
	Intr	41203	41259	+	0
	Intr	41359	41437	+	0
	Intr	41521	41681	+	0
45	Intr	41788	41915	+	0
	Intr	42010	42147	+	0
	Intr	42474	42566	+	0
	Intr	42653	42773	+	0
	Term	42877	43233	+	0

>4803950 /7133

	len =	1106	nex =	2	
55	Term	15958	15293	-	0
	Init	16398	16136	-	0

>4809270 /21854

60	len =	1810	nex =	6	
----	-------	------	-------	---	--

1545

	Init	26325	26534	+	0
	Intr	26761	26841	+	0
5		26913	27026	+	0
	Intr	27129	27407	+	0
	Intr	27488	27601	+	0
	Term	27698	27770	+	0
10	>4809270	/23558			
	len =	674	nex =	1	
	Sngl	40607	41280	+	0
15	>4809270	/36972			
	len =	2847	nex =	11	
20	Init	64059	64119	+	0
	Intr	64238	64297	+	0
	Intr	64402	64477	+	0
	Intr	64578	64743	+	0
	Intr	64827	65082	+	0
25	Intr	65192	65307	+	0
	Intr	65390	65610	+	0
	Intr	65694	65728	+	0
	Intr	65818	65917	+	0
	Intr	66016	66118	+	0
30	Term	66195	66599	+	0
	>4809270	/20668			
	len =	1075	nex =	5	
35	Init	65522	65610	+	0
	Intr	65694	65728	+	0
	Intr	65818	65917	+	0
	Intr	66016	66118	+	0
40	Term	66195	66596	+	0
	>4809271	/21905			
	len =	1099	nex =	2	
45	Term	12979	12614	-	0
	Init	13712	13449	-	0
	>4809271	/22359			
50	len =	1095	nex =	1	
	Sngl	13712	12618	-	0
55	>4809271	/848			
	len =	760	nex =	1	
	Sngl	33077	32318	-	0
60	>4809271	/102190			

1546

	len =	743	nex =	1	
5	Sngl	33077	32335	-	0
	>4809271	/21181			
	len =	344	nex =	1	
10	Sngl	8636	8293	-	0
	>4809271	/6687			
	len =	1406	nex =	1	
15	Sngl	8731	7326	-	0
	>4809294	/8411			
20	len =	1440	nex =	4	
	Init	14573	14846	+	0
	Intr	15067	15169	+	0
	Intr	15295	15456	+	0
25	Term	15734	16012	+	0
	>4809294	/8986			
	len =	2265	nex =	7	
30	Init	21367	21498	+	0
	Intr	21594	21784	+	0
	Intr	21897	22034	+	0
	Intr	22122	22258	+	0
35	Intr	22349	22514	+	0
	Intr	22682	22921	+	0
	Term	23003	23631	+	0
	>4809294	/33512			
40	len =	1150	nex =	2	
	Term	30857	30449	-	0
	Init	31590	31115	-	0
45	>4809295	/3676			
	len =	2083	nex =	6	
50	Term	53797	53671	-	0
	Intr	53990	53913	-	0
	Intr	54177	54096	-	0
	Intr	54877	54684	-	0
	Intr	55272	55181	-	0
55	Init	55753	55348	-	0
	>4809296	/30187			
60	len =	1390	nex =	3	

1547

Term	20449	20131	-	0
Intr	20937	20890	-	0
Init	21511	21158	-	0

5 >4809296 /3071

len = 1092 nex = 4

Init	411	493	+	0
Intr	704	768	+	0
Intr	971	1041	+	0
Term	1133	1359	+	0

>4809296 /34078

15 len = 1150 nex = 4

Init	411	493	+	0
Intr	704	768	+	0
Intr	971	1041	+	0
Term	1133	1416	+	0

>4835223 /36702

25 len = 2050 nex = 8

Term	28902	28573	-	0
Intr	29030	28977	-	0
Intr	29207	29113	-	0
Intr	29404	29307	-	0
Intr	29625	29483	-	0
Intr	29839	29708	-	0
Intr	30116	29988	-	0
Init	30618	30194	-	0

>4835223 /3286

len = 562 nex = 1

40 Sngl 40069 39514 - 0

>4835223 /8864

len = 1585 nex = 4

Term	42999	42518	-	0
Intr	43254	43184	-	0
Intr	43459	43352	-	0
Init	44102	43814	-	0

50 >4835223 /153949

len = 1900 nex = 6

Init	50935	51051	+	0
Intr	51154	51196	+	0
Intr	51320	51381	+	0
Intr	51479	51575	+	0
Intr	52119	52189	+	0
Term	52348	52439	+	0

1548

>4835223 /27213

5	len =	1570	nex =	2	
	Init	59627	60267	+	0
	Term	60638	61187	+	0

>4835223 /732

10	len =	1474	nex =	1	
	Sngl	62121	62953	+	0

15 >4835223 /20695

len =	1163	nex =	0
-------	------	-------	---

20 >4835223 /23377

len =	790	nex =	1	
Sngl	65775	64987	-	0

25 >4835223 /10044

len =	495	nex =	1	
Sngl	66699	67193	+	0

30 >4835223 /32364

len =	1750	nex =	1	
Sngl	78553	80295	+	0

>4835223 /5193

40	len =	1348	nex =	8	
	Term	87820	87718	-	0
	Intr	87968	87907	-	0
	Intr	88132	88060	-	0
	Intr	88361	88219	-	0
45	Intr	88494	88447	-	0
	Intr	88643	88569	-	0
	Intr	88863	88780	-	0
	Init	89065	88950	-	0

50 >4835773 /1989

len =	690	nex =	1	
Sngl	20133	20822	+	0

55 >4835773 /34385

len =	2830	nex =	12	
Init	2024	2254	+	0

2025-09-10 09:16:00

1549

	Intr	2371	2450	+	0
	Intr	2617	2733	+	0
	Intr	2833	2936	+	0
	Intr	3023	3089	+	0
5	Intr	3363	3496	+	0
	Intr	3582	3630	+	0
	Intr	3767	3838	+	0
	Intr	3929	4030	+	0
	Intr	4116	4183	+	0
10	Intr	4369	4521	+	0
	Term	4602	4852	+	0
>4835773 /6281					
15	len =	1946	nex =	7	
	Init	42427	42503	+	0
	Intr	42604	42707	+	0
	Intr	42803	42903	+	0
20	Intr	42991	43094	+	0
	Intr	43278	43715	+	0
	Intr	43797	43928	+	0
	Term	43997	44372	+	0
25	>4835773 /28214				
	len =	1757	nex =	3	
	Term	48583	47836	-	0
30	Intr	48864	48677	-	0
	Init	49592	48965	-	0
>4835773 /21943					
35	len =	1870	nex =	4	
	Init	5082	5505	+	0
	Intr	5583	5729	+	0
	Intr	5834	6043	+	0
40	Term	6473	6945	+	0
>4836442 /1137					
	len =	730	nex =	1	
45	Sngl	5764	6489	+	0
>4836906 /93312					
50	len =	1831	nex =	4	
	Term	26269	26014	-	0
	Intr	26453	26358	-	0
	Intr	26664	26531	-	0
55	Init	27844	26730	-	0
>4836906 /29867					
60	len =	831	nex =	1	

1550

	Sngl	64559	63729	-	0
	>4836906	/14139			
5	len =	1570	nex =	2	
	Term	80467	80066	-	0
	Init	80887	80566	-	0
10	>4836906	/5665			
	len =	898	nex =	2	
	Init	86401	86906	+	0
15	Term	87082	87298	+	0
	>4850281	/4143			
	len =	436	nex =	1	
20	Sngl	107876	108311	+	0
	>4850409	/12642			
25	len =	583	nex =	2	
	Term	104024	103684	-	0
	Init	104266	104150	-	0
30	>4850409	/20585			
	len =	730	nex =	2	
	Init	106425	106547	+	0
35	Term	106829	107153	+	0
	>4850409	/28536			
	len =	296	nex =	1	
40	Sngl	106829	107124	+	0
	>4850409	/39306			
45	len =	2314	nex =	5	
	Init	14095	14194	+	0
	Intr	14334	14483	+	0
	Intr	14590	14863	+	0
50	Intr	14895	14982	+	0
	Term	15124	15585	+	0
	>4850409	/3062			
55	len =	2530	nex =	7	
	Term	18950	18748	-	0
	Intr	19135	19019	-	0
	Intr	19377	19225	-	0
60	Intr	19637	19467	-	0

1551

	Intr	19900	19721	-	0
	Intr	20082	19984	-	0
	Init	20494	20407	-	0
5	>4850409	/4791			
	len =	736	nex =	0	
10	>4850409	/27686			
	len =	716	nex =	1	
	Sngl	23412	24127	+	0
15	>4850409	/30722			
	len =	1677	nex =	4	
20	Init	34646	34737	+	0
	Intr	34898	35330	+	0
	Intr	35419	35472	+	0
	Term	35528	35913	+	0
25	>4850409	/94765			
	len =	235	nex =	1	
	Sngl	35586	35820	+	0
30	>4850409	/40551			
	len =	1848	nex =	4	
35	Term	62292	61937	-	0
	Intr	62986	62451	-	0
	Intr	63104	63016	-	0
	Init	63784	63510	-	0
40	>4850409	/112273			
	len =	850	nex =	2	
	Term	66692	66564	-	0
	Init	66880	66784	-	0
45	>4850409	/5724			
	len =	1295	nex =	2	
50	Term	66692	66532	-	0
	Init	66880	66784	-	0
	>4850409	/2548			
55	len =	692	nex =	1	
	Sngl	67705	68396	+	0
60	>4850409	/8711			

1552

	len =	708	nex =	1	
	Sngl	67705	68412	+	0
5	>4850409	/2535			
	len =	598	nex =	1	
	Sngl	69019	69616	+	0
10	>4850409	/30316			
	len =	790	nex =	1	
15	Sngl	76323	75925	-	0
	>4850409	/7539			
	len =	791	nex =	1	
20	Sngl	76452	75925	-	0
	>4850409	/4289			
25	len =	1643	nex =	1	
	Sngl	90019	91371	+	0
	>4850409	/27754			
30	len =	1431	nex =	1	
	Sngl	89946	91376	+	0
35	>4850411	/21257			
	len =	430	nex =	1	
	Sngl	14482	14057	-	0
40	>4874280	/12402			
	len =	1492	nex =	4	
45	Init	100913	101010	+	0
	Intr	101624	101737	+	0
	Intr	101836	101915	+	0
	Term	102002	102400	+	0
50	>4874280	/13543			
	len =	670	nex =	1	
	Sngl	110726	110057	-	0
55	>4874280	/29810			
	len =	1717	nex =	7	
60	Term	115803	115457	-	0

1553

	Intr	116038	115976	-	0
	Intr	116268	116158	-	0
	Intr	116431	116349	-	0
	Intr	116759	116512	-	0
5	Intr	116980	116882	-	0
	Init	117173	117074	-	0
	>4874280		/14469		
10	len =	146	nex =	1	
	Sngl	25949	25804	-	0
	>4874280		/1871		
15	len =	1210	nex =	1	
	Sngl	37446	36737	-	0
20	>4874280		/25201		
	len =	1044	nex =	6	
25	Term	43874	43821	-	0
	Intr	44112	43996	-	0
	Intr	44365	44198	-	0
	Intr	44512	44447	-	0
	Intr	44684	44589	-	0
	Init	44864	44764	-	0
30	>4874280		/10068		
	len =	740	nex =	1	
35	Sngl	57061	56322	-	0
	>4874280		/11912		
40	len =	772	nex =	4	
	Term	62694	62514	-	0
	Intr	62885	62784	-	0
	Intr	63127	62992	-	0
	Init	63285	63202	-	0
45	>4874280		/30239		
	len =	758	nex =	1	
50	Sngl	63635	64392	+	0
	>4874280		/22166		
55	len =	1400	nex =	4	
	Term	64736	64526	-	0
	Intr	65314	65273	-	0
	Intr	65723	65668	-	0
	Init	65925	65828	-	0
60					

2025 RELEASE UNDER E.O. 14176

1554

>4874280 /8886

```
len =      271      nex =      1
```

5	Snq1	83252	82982	-	0
---	------	-------	-------	---	---

>4874280 /21847

```
len = 1253    nex = 2
```

```

10      Init      88169      88958      +      0

```

Term	89032	89421	+	0
------	-------	-------	---	---

>4874280 /19135

```

15      len =      5291      nex =          3

```

Term	86468	86081	-	0
------	-------	-------	---	---

Intr	87158	86630	-	0
------	-------	-------	---	---

Init	91371	91235	-	0
------	-------	-------	---	---

>4874280 /16209

```
len = 1870    nex = 2
```

Term	90809	89582	-	0
------	-------	-------	---	---

Init	91445	91235	-	0
------	-------	-------	---	---

>4874280 /9177

```
len = 1603    nex = 5
```

Term	92357	92087	-	0
------	-------	-------	---	---

Intr	92498	92429	-	0
------	-------	-------	---	---

Intr	92677	92587	-	0
------	-------	-------	---	---

Intr	93017	92768	-	0
------	-------	-------	---	---

Init	93689	93526	-	0
------	-------	-------	---	---

>4874280 /17356

```

40      len =      1051      nex =      1

```

Sngl 97810 98457 + 0

45 >4874280 /121746

```
len = 752    nex = 1
```

Snq1 97770 98521 + 0

50

>4874281 /18634

```
len = 992    nex = 1
```

55	Snq1	28766	29757	+	
----	------	-------	-------	---	--

>4878038 /33093

```

60      len =      1838      nex =

```

1555

	Init	17045	17083	+	0
	Intr	17496	17715	+	0
	Intr	17935	18104	+	0
	Term	18283	18716	+	0
5	>4878038 /12279				
	len =	1657	nex =	4	
10	Init	17042	17083	+	0
	Intr	17496	17715	+	0
	Intr	17935	18104	+	0
	Term	18283	18698	+	0
15	>4878038 /41419				
	len =	101	nex =	1	
20	Sngl	1922	1822	-	0
	>4878038 /103167				
	len =	827	nex =	3	
25	Term	22122	21766	-	0
	Intr	22412	22340	-	0
	Init	22592	22509	-	0
30	>4878038 /12044				
	len =	832	nex =	3	
	Term	22122	21767	-	0
	Intr	22412	22340	-	0
35	Init	22598	22509	-	0
	>4878038 /7414				
40	len =	850	nex =	3	
	Term	22122	21767	-	0
	Intr	22412	22340	-	0
	Init	22611	22509	-	0
45	>4878038 /39977				
	len =	915	nex =	3	
50	Term	22122	21728	-	0
	Intr	22412	22340	-	0
	Init	22642	22509	-	0
	>4878038 /18748				
55	len =	635	nex =	3	
	Term	22122	22008	-	0
	Intr	22412	22340	-	0
	Init	22642	22509	-	0
60					

1556

>4878038 /34623

```
len =      911      nex =      3
```

5	Term	22122	21766	-	0
	Intr	22412	22340	-	0
	Init	22676	22509	-	0

>4878038 /35553

```
10      len = 1034      nex = 3
```

	Init	23065	23187	+	0
	Intr	23268	23447	+	0
15	Term	23838	24098	+	0

>4878038 /40238

```
len = 1210    nex = 1
```

Snq1	2380	1177	-	0
------	------	------	---	---

>4878038 /107233

```
25      len =      643      nex =      4
```

	Term	48867	48732	-	0
	Intr	49015	48983	-	0
	Intr	49222	49136	-	0
30	Init	49374	49316	-	0

>4878038 /35098

```
len = 2170 nex = 3
```

Init	53312	53618	+	0
Intr	53769	54355	+	0
Term	55015	55476	+	0

40 >4878038 /6413

```
len = 3951    nex =
```

	Init	56661	56949	+	0
45	Intr	58951	59082	+	0
	Intr	59203	59289	+	0
	Intr	59403	59453	+	0
	Term	59530	60611	+	0

50 >4878038 /119807

```
len = 1461    nex =
```

	Init	59122	59289	+
55	Intr	59403	59453	+
	Term	60354	60582	+

>4878038 /34128

```
60      len =      1012      nex =
```

1557

	Init	65143	65373	+	0
	Intr	65735	65837	+	0
	Term	66104	66154	+	0
5	>4878038 /123625				
	len =	897	nex =	2	
10	Init	69908	70156	+	0
	Term	70242	70804	+	0
	>4878038 /150460				
15	len =	313	nex =	1	
	Sngl	75227	74915	-	0
	>4878039 /27589				
20	len =	1529	nex =	2	
	Init	21799	21871	+	0
	Term	22351	22674	+	0
25	>4878039 /31701				
	len =	1491	nex =	2	
30	Init	21799	21871	+	0
	Term	22351	22658	+	0
	>4878039 /22285				
35	len =	1224	nex =	1	
	Sngl	23812	23609	-	0
	>4878039 /7311				
40	len =	1244	nex =	1	
	Sngl	23835	23609	-	0
45	>4878039 /13410				
	len =	1243	nex =	0	
	>4878039 /11384				
50	len =	534	nex =	1	
	Sngl	41117	40584	-	0
55	>4878039 /21735				
	len =	2391	nex =	6	
	Init	6076	6252	+	0
60	Intr	6656	6778	+	0

2025 RELEASE UNDER E.O. 14176

1558

	Intr	6934	7050	+	0
	Intr	7141	7336	+	0
	Intr	7584	7698	+	0
	Term	7795	7959	+	0
5	>4878039 /153670				
	len =	2383	nex =	6	
10	Init	6084	6252	+	0
	Intr	6656	6778	+	0
	Intr	6934	7050	+	0
	Intr	7141	7336	+	0
	Intr	7584	7698	+	0
15	Term	7795	7959	+	0
	>4878039 /118900				
	len =	2308	nex =	6	
20	Init	6091	6252	+	0
	Intr	6656	6778	+	0
	Intr	6934	7050	+	0
	Intr	7141	7336	+	0
25	Intr	7584	7698	+	0
	Term	7795	7959	+	0
	>4878039 /17026				
30	len =	156	nex =	1	
	Sngl	61886	61731	-	0
	>4878039 /24663				
35	len =	334	nex =	1	
	Sngl	61970	61637	-	0
40	>4878039 /15897				
	len =	458	nex =	1	
	Sngl	62053	61601	-	0
45	>4878039 /9325				
	len =	1366	nex =	4	
50	Term	62075	61726	-	0
	Intr	62381	62331	-	0
	Intr	62704	62472	-	0
	Init	63091	62815	-	0
55	>4878039 /33019				
	len =	1400	nex =	4	
	Term	62075	61700	-	0
60	Intr	62381	62331	-	0

1559

Intr	62704	62472	-	0
Init	63099	62815	-	0

>4878039 /14650

5	len =	190	nex =	1	
	Sngl	63116	62933	-	0

10 >4878039 /41423

len =	1419	nex =	0
-------	------	-------	---

15 >4878039 /33922

len =	1522	nex =	4
-------	------	-------	---

	Term	62075	61601	-	0
	Intr	62381	62331	-	0
20	Intr	62704	62472	-	0
	Init	63122	62815	-	0

>4878039 /8579

25	len =	813	nex =	3	
	Init	72885	73192	+	0
	Intr	73271	73362	+	0
	Term	73447	73688	+	0

30 >4878039 /38014

len =	1795	nex =	3
-------	------	-------	---

35	Term	81714	81447	-	0
	Intr	81951	81802	-	0
	Init	82973	82025	-	0

>4878039 /6888

40	len =	1128	nex =	2	
	Term	84567	84273	-	0
	Init	85400	84860	-	0

45 >4883587 /206485

len =	538	nex =	1
-------	-----	-------	---

50 Sngl 113867 113330 - 0

>4883587 /8904

len =	1657	nex =	4
-------	------	-------	---

55	Init	1398	1648	+	0
	Intr	2403	2594	+	0
	Intr	2697	2865	+	0
	Term	2973	3054	+	0

60

1560

>4883587 /99140

```
len =      612      nex =      1
```

5	Snq1	81836	81225	-	0
---	------	-------	-------	---	---

>4883587 /35520

```
len = 1890    nex = 1
```

10					
	Snq1	83089	81200	-	0

>4883588 /30158

```
15      len =      1058      nex =          3
```

Init	78085	78219	+	0
------	-------	-------	---	---

Intr	78340	78388	+	0
------	-------	-------	---	---

Term	78702	79142	+	0
------	-------	-------	---	---

20
 >4883588 /42159

```
len =      938      nex =      3
```

25	Init	78151	78219	+	0
----	------	-------	-------	---	---

Intr	78340	78388	+	0
------	-------	-------	---	---

Term	78702	79088	+	0
------	-------	-------	---	---

>4883588 /117103

```
len = 956    nex = 2
```

Unit	78187	78219	+	
------	-------	-------	---	--

Term	78340	79142	+	
------	-------	-------	---	--

35
 >4883588 /1296

```
len = 2950    nex = 10
```

40	Term	79984	79379	-
----	------	-------	-------	---

Intr	80178	80092	-
------	-------	-------	---

Intr	80331	80272	-
------	-------	-------	---

Intr	80484	80416	-
------	-------	-------	---

Intr	80729	80573	-
------	-------	-------	---

45	Intr	80924	80848	-
----	------	-------	-------	---

Intr	81328	81283	-
------	-------	-------	---

Intr	81508	81422	-
------	-------	-------	---

Intr	81730	81584	-
------	-------	-------	---

Init	82321	81955	-
------	-------	-------	---

50 >4883589 /3947

```
len = 910    nex =
```

55	Snq1	27167	26259	-
----	------	-------	-------	---

>4883595 /2717

```
len = 1305    nex =
```

60

1561

	Init	13666	13838	+	0
	Intr	14082	14123	+	0
	Intr	14233	14403	+	0
	Intr	14555	14641	+	0
5	Term	14739	14970	+	0

>4883595 /28643

10	len =	1274	nex =	5	
	Init	13690	13838	+	0
	Intr	14082	14123	+	0
	Intr	14233	14403	+	0
	Intr	14555	14641	+	0
15	Term	14739	14963	+	0

>4883595 /27627

20	len =	1278	nex =	5	
	Init	13690	13838	+	0
	Intr	14082	14123	+	0
	Intr	14233	14403	+	0
	Intr	14555	14641	+	0
25	Term	14739	14967	+	0

>4883595 /121982

30	len =	1275	nex =	5	
	Init	13695	13838	+	0
	Intr	14082	14123	+	0
	Intr	14233	14403	+	0
	Intr	14555	14641	+	0
35	Term	14739	14969	+	0

>4883595 /25350

40	len =	1004	nex =	5	
	Init	13767	13838	+	0
	Intr	14082	14123	+	0
	Intr	14233	14403	+	0
	Intr	14555	14641	+	0
45	Term	14739	14764	+	0

>4883595 /32137

50	len =	1896	nex =	5	
	Init	38588	38819	+	0
	Intr	39455	39511	+	0
	Intr	39688	39878	+	0
	Intr	39967	40071	+	0
55	Term	40162	40483	+	0

>4883595 /37242

60	len =	1971	nex =	5	
----	-------	------	-------	---	--

00100-0220000

1562

	Init	38588	38819	+	0
	Intr	39455	39511	+	0
	Intr	39688	39878	+	0
	Intr	39967	40071	+	0
5	Term	40162	40558	+	0
>4883595 /20760					
10	len =	1990	nex =	5	
	Init	38780	38819	+	0
	Intr	39455	39511	+	0
	Intr	39688	39878	+	0
	Intr	39967	40071	+	0
15	Term	40162	40576	+	0
>4883595 /34869					
20	len =	899	nex =	1	
	Sngl	38590	38819	+	0
>4883595 /25104					
25	len =	372	nex =	1	
	Sngl	40223	40578	+	0
>4883595 /5109					
30	len =	1376	nex =	4	
	Term	66464	66136	-	0
	Intr	66734	66627	-	0
35	Intr	66906	66815	-	0
	Init	67219	67018	-	0
>4883595 /113853					
40	len =	912	nex =	3	
	Init	67833	68023	+	0
	Intr	68119	68397	+	0
	Term	68518	68744	+	0
45	>4883595 /27387				
	len =	1095	nex =	1	
50	Sngl	97707	96779	-	0
>4883598 /112395					
55	len =	102	nex =	1	
	Sngl	22321	22422	+	0
>4883598 /27175					
60	len =	1292	nex =	6	

1563

	Term	53471	53305	-	0
	Intr	53639	53577	-	0
	Intr	53815	53735	-	0
5	Intr	53963	53919	-	0
	Intr	54304	54215	-	0
	Init	54443	54387	-	0
	>4883598		/31249		
10	len =	1066	nex =	3	
	Init	75625	75768	+	0
	Intr	75862	75941	+	0
15	Term	76089	76422	+	0
	>4883599		/41933		
	len =	2470	nex =	6	
20	Init	107705	108086	+	0
	Intr	108321	108417	+	0
	Intr	108494	108609	+	0
	Intr	109315	109359	+	0
25	Intr	109639	109774	+	0
	Term	109859	110173	+	0
	>4883599		/19094		
30	len =	1768	nex =	1	
	Sngl	118037	116270	-	0
	>4883599		/11488		
35	len =	696	nex =	1	
	Sngl	119284	118589	-	0
40	>4883599		/97088		
	len =	1795	nex =	4	
	Term	11068	10803	-	0
45	Intr	11765	11697	-	0
	Intr	12260	12150	-	0
	Init	12597	12343	-	0
	>4883599		/7831		
50	len =	2010	nex =	4	
	Term	34414	33824	-	0
	Intr	35311	35094	-	0
55	Intr	35494	35394	-	0
	Init	35833	35596	-	0
	>4883599		/13320		
60	len =	2015	nex =	1	

1564

	Sngl	42074	40567	-	0
	>4883599 /29157				
5	len =	1707	nex =	6	
	Term	43450	43184	-	0
	Intr	43605	43533	-	0
10	Intr	43874	43750	-	0
	Intr	44036	43965	-	0
	Intr	44322	44225	-	0
	Init	44890	44558	-	0
15	>4883599 /16221				
	len =	1270	nex =	5	
	Term	50525	50232	-	0
20	Intr	50711	50610	-	0
	Intr	50909	50793	-	0
	Intr	51085	50985	-	0
	Init	51498	51268	-	0
25	>4883599 /2416				
	len =	1352	nex =	4	
	Term	52955	52650	-	0
30	Intr	53295	53025	-	0
	Intr	53806	53522	-	0
	Init	54001	53894	-	0
	>4883599 /36633				
35	len =	1133	nex =	2	
	Init	6605	7052	+	0
40	Term	7144	7737	+	0
	>4883599 /119288				
	len =	646	nex =	2	
45	Init	6611	7052	+	0
	Term	7144	7256	+	0
	>4883599 /12558				
50	len =	2034	nex =	3	
	Init	66457	66727	+	0
	Intr	66744	66880	+	0
	Term	66958	68472	+	0
55	>4883599 /12259				
	len =	212	nex =	1	
60	Sngl	67175	67386	+	0

1565

>4883599 /39855

5	len =	1599	nex =	4	
	Term	71283	71095	-	0
	Intr	71609	71547	-	0
	Intr	71898	71722	-	0
	Init	72208	72146	-	0

>4883599 /19570

	len =	1714	nex =	3	
15	Init	76320	76853	+	0
	Intr	77103	77181	+	0
	Term	77311	78033	+	0

>4883599 /3766

20	len =	631	nex =	2	
	Term	98968	98595	-	0
	Init	99225	99034	-	0

>4884020 /8077

	len =	1751	nex =	3	
30	Term	15439	14937	-	0
	Intr	15776	15562	-	0
	Init	16497	16374	-	0

>4884020 /5161

35	len =	2037	nex =	8	
	Term	22090	22010	-	0
	Intr	22315	22187	-	0
40	Intr	22607	22491	-	0
	Intr	22782	22718	-	0
	Intr	22906	22867	-	0
	Intr	23079	23017	-	0
	Intr	23444	23183	-	0
45	Init	23627	23517	-	0

>4884020 /27539

	len =	1335	nex =	4	
50	Init	34674	35251	+	0
	Intr	35340	35450	+	0
	Intr	35592	35705	+	0
	Term	35914	36008	+	0

>4884021 /2419

	len =	1136	nex =	2	
60	Init	33897	34058	+	0

1566

	Term	34433	35032	+	0
	>4884021 /19314				
5	len =	2613	nex =	2	
	Init	3435	3738	+	0
	Term	3868	6047	+	0
10	>4884022 /2982				
	len =	2050	nex =	8	
	Init	25501	25640	+	0
15	Intr	25886	26005	+	0
	Intr	26146	26292	+	0
	Intr	26381	26476	+	0
	Intr	26560	26655	+	0
	Intr	26927	27044	+	0
20	Intr	27132	27204	+	0
	Term	27301	27548	+	0
	>4884022 /40445				
25	len =	3512	nex =	12	
	Term	28180	27923	-	0
	Intr	28424	28270	-	0
	Intr	28550	28502	-	0
30	Intr	28909	28653	-	0
	Intr	29159	29030	-	0
	Intr	29337	29269	-	0
	Intr	29632	29579	-	0
	Intr	29897	29739	-	0
35	Intr	30278	30141	-	0
	Intr	30434	30372	-	0
	Intr	30577	30538	-	0
	Init	30707	30661	-	0
40	>4884023 /94438				
	len =	1247	nex =	3	
	Init	32053	32253	+	0
45	Intr	32590	32658	+	0
	Term	32956	33299	+	0
	>4884023 /40485				
50	len =	2060	nex =	7	
	Init	76282	76657	+	0
	Intr	76812	76874	+	0
	Intr	77011	77127	+	0
55	Intr	77202	77448	+	0
	Intr	77547	77705	+	0
	Intr	77780	77931	+	0
	Term	78015	78341	+	0
60	>4884023 /110908				

2025-09-23 14:00:00

1567

len = 372 nex = 0

>4886265 /16125

5 len = 1665 nex = 4

Init 15167 15326 + 0

Intr 15839 15978 + 0

10 Intr 16219 16344 + 0

Term 16429 16831 + 0

>4886265 /39625

15 len = 372 nex = 1

Sngl 45373 45744 + 0

>4886265 /11782

20 len = 910 nex = 1

Sngl 58085 58990 + 0

>4886265 /92178

25 len = 910 nex = 1

Sngl 60596 59692 - 0

30 >4886265 /7969

len = 1105 nex = 5

35 Term 6613 6546 - 0

Intr 6828 6699 - 0

Intr 7010 6925 - 0

Intr 7452 7407 - 0

Init 7643 7530 - 0

40 >4886265 /9716

len = 1646 nex = 6

45 Term 6613 6377 - 0

Intr 6828 6699 - 0

Intr 7010 6925 - 0

Intr 7452 7407 - 0

Intr 7649 7530 - 0

50 Init 8022 7921 - 0

>4886265 /37248

len = 1450 nex = 2

55 Init 84108 84785 + 0

Term 84958 85556 + 0

>4886265 /14246

60

1568

	len =	1696	nex =	4	
	Term	88532	88155	-	0
	Intr	88912	88604	-	0
5	Intr	89088	88987	-	0
	Init	89850	89191	-	0
>4887257		/145295			
10	len =	1695	nex =	4	
	Term	10323	10083	-	0
	Intr	10562	10518	-	0
	Intr	10742	10642	-	0
15	Init	11039	10834	-	0
>4887257		/22508			
20	len =	1006	nex =	4	
	Term	23643	23309	-	0
	Intr	23922	23824	-	0
	Intr	24104	24023	-	0
	Init	24314	24189	-	0
25	>4887257		/143886		
	len =	850	nex =	3	
30	Init	69630	69796	+	0
	Intr	69886	69951	+	0
	Term	70176	70475	+	0
>4887257		/39490			
35	len =	2950	nex =	12	
	Term	74958	74703	-	0
	Intr	75114	75047	-	0
40	Intr	75239	75176	-	0
	Intr	75402	75321	-	0
	Intr	75574	75516	-	0
	Intr	75705	75646	-	0
	Intr	75982	75820	-	0
45	Intr	76205	76081	-	0
	Intr	76563	76501	-	0
	Intr	76761	76702	-	0
	Intr	77059	76880	-	0
	Init	77646	77488	-	0
50	>4887257		/8633		
	len =	2801	nex =	10	
55	Init	78250	78326	+	0
	Intr	78410	78474	+	0
	Intr	78598	78644	+	0
	Intr	78737	78812	+	0
	Intr	78900	78967	+	0
60	Intr	79275	79344	+	0

1569

	Intr	79628	79713	+	0
	Intr	79880	79942	+	0
	Intr	80237	80302	+	0
	Term	80400	80609	+	0
5	>4887257 /109912				
	len =	1124	nex =	3	
10	Init	94040	94098	+	0
	Intr	94128	94314	+	0
	Term	94452	95163	+	0
	>4887737 /2004				
15	len =	675	nex =	1	
	Sngl	7907	8581	+	0
20	>4887737 /1505				
	len =	673	nex =	1	
	Sngl	7909	8581	+	0
25	>4887737 /31675				
	len =	599	nex =	1	
30	Sngl	7913	8511	+	0
	>4887737 /106951				
	len =	677	nex =	1	
35	Sngl	7913	8589	+	0
	>4887737 /108174				
40	len =	1630	nex =	2	
	Term	85528	84876	-	0
	Init	86500	86083	-	0
45	>4887737 /33557				
	len =	1788	nex =	2	
	Term	85528	84753	-	0
50	Init	86540	86083	-	0
	>4887738 /23872				
	len =	764	nex =	1	
55	Sngl	20581	19818	-	0
	>4887738 /119765				
60	len =	3139	nex =	7	

2025 RELEASE UNDER E.O. 14176

1570

		Init	3396	4329	+	0
		Intr	4533	4587	+	0
		Intr	5001	5039	+	0
5		Intr	5122	5185	+	0
		Intr	5549	5576	+	0
		Intr	5741	5832	+	0
		Term	6013	6534	+	0
10	>4887738	/114822				
	len =	3136	nex =	7		
		Init	3399	4329	+	0
15		Intr	4533	4587	+	0
		Intr	5001	5039	+	0
		Intr	5122	5185	+	0
		Intr	5549	5576	+	0
		Intr	5741	5832	+	0
20		Term	6013	6534	+	0
	>4887738	/28177				
	len =	1570	nex =	3		
25		Term	43839	43444	-	0
		Intr	44038	43918	-	0
		Init	45006	44725	-	0
30	>4887738	/6612				
	len =	2489	nex =	4		
		Init	61698	61850	+	0
35		Intr	61949	62099	+	0
		Intr	62198	62262	+	0
		Term	62746	63244	+	0
	>4887738	/29992				
40		len =	926	nex =	2	
		Term	78325	77774	-	0
45		Init	78699	78401	-	0
	>4887738	/146828				
	len =	507	nex =	1		
50	Sngl	83673	83167	-		0
	>4887740	/33192				
	len =	2290	nex =	5		
55		Init	199	824	+	0
		Intr	922	1089	+	0
		Intr	1185	1376	+	0
		Intr	1551	1619	+	0
60		Term	1985	2290	+	0

1571

>4887740 /23467

```

5      len =      521      nex =      2
      Init  46831  46988      +      0
      Term  47080  47203      +      0

```

>4887740 /12467

```

10      len =      651    nex =      1
      Snq1  5943    5293    -      0

```

15 >4895147 /124537

	len =	1961	nex =	7	
20	Init	15249	15368	+	0
	Intr	15727	15850	+	0
	Intr	15938	15966	+	0
	Intr	16050	16115	+	0
	Intr	16202	16309	+	0
25	Intr	16399	16464	+	0
	Term	16795	16876	+	0

>4895147 /21739

```

30      len =      1469      nex =          4
      Term 50700  50490      -          0
      Intr 51095  50898      -          0
      Intr 51286  51185      -          0
      Init 51958  51514      -          0

```

35
 >4895147 /21920

```

len =      444    nex =      1
40      Snq1 53076 52638      -      0

```

>4895147 /37658

```

45      len =      1662      nex =          3
      Term 53123  52698      -      0
      Intr 53993  53216      -      0
      Init 54359  54077      -      0

```

50 >4895147 /24367

```

len =      993  nex =      3
55  Term 55478 55257 - 0
    Intr 55772 55560 - 0
    Init 56104 56024 - 0

```

>4895147 /20633

```
60      len = 1066      nex = 3
```

1572

	Term	55478	55258	-	0
	Intr	55772	55560	-	0
	Init	56104	56024	-	0
5	>4895147	/39533			
	len =	730	nex =	2	
10	Term	55772	55601	-	0
	Init	56104	56024	-	0
	>4895147	/4458			
15	len =	740	nex =	2	
	Term	55772	55598	-	0
	Init	56104	56024	-	0
20	>4895147	/20676			
	len =	1041	nex =	2	
	Term	55478	55286	-	0
25	Init	56104	56024	-	0
	>4895147	/524			
	len =	1101	nex =	1	
30	Sngl	56104	56024	-	0
	>4895147	/6803			
35	len =	979	nex =	3	
	Term	55478	55354	-	0
	Intr	55772	55560	-	0
40	Init	56104	56024	-	0
	>4895147	/15992			
	len =	1018	nex =	3	
45	Term	55478	55318	-	0
	Intr	55772	55560	-	0
	Init	56104	56024	-	0
	>4895147	/24607			
50	len =	550	nex =	1	
	Sngl	68923	68382	-	0
55	>4895147	/43068			
	len =	2470	nex =	10	
	Term	70358	70271	-	0
60	Intr	70597	70472	-	0

1573

	Intr	70793	70704	-	0
	Intr	71005	70880	-	0
	Intr	71301	71248	-	0
	Intr	71483	71403	-	0
5	Intr	71698	71591	-	0
	Intr	71890	71819	-	0
	Intr	72038	71976	-	0
	Init	72731	72465	-	0
10	>4895147 /20749				
	len =	2192	nex =	6	
	Init	7453	7722	+	0
15	Intr	7862	7924	+	0
	Intr	8068	8184	+	0
	Intr	8583	8988	+	0
	Intr	9110	9261	+	0
	Term	9343	9644	+	0
20	>4895147 /28975				
	len =	1253	nex =	4	
25	Init	77491	77810	+	0
	Intr	78228	78304	+	0
	Intr	78428	78536	+	0
	Term	78629	78743	+	0
30	>4895147 /98397				
	len =	730	nex =	1	
35	Sngl	79391	79138	-	0
	>4895147 /37526				
	len =	1066	nex =	1	
40	Sngl	80755	81820	+	0
	>4895148 /14553				
	len =	1704	nex =	2	
45	Term	39478	38899	-	0
	Init	40602	39746	-	0
	>4895148 /114523				
50	len =	310	nex =	1	
	Sngl	55690	55388	-	0
55	>4895148 /40329				
	len =	1066	nex =	2	
	Term	55806	55404	-	0
60	Init	56469	56070	-	0

ORIGINAL - 01/25/50

1574

5 >4895148 /109156
 len = 480 nex = 1
 Sngl 57879 57400 - 0
 10 >4895149 /37283
 len = 1572 nex = 2
 Term 20933 20082 - 0
 Init 21653 21105 - 0
 15 >4895167 /18533
 len = 892 nex = 3
 20 Init 43380 43470 + 0
 Intr 43555 43730 + 0
 Term 43825 44256 + 0
 25 >4895176 /34547
 len = 2128 nex = 6
 Term 104785 104678 - 0
 Intr 105146 104871 - 0
 Intr 105483 105396 - 0
 30 Intr 105659 105583 - 0
 Intr 106058 106008 - 0
 Init 106328 106140 - 0
 35 >4895176 /6630
 len = 1966 nex = 6
 Term 111441 111255 - 0
 Intr 112230 111955 - 0
 40 Intr 112394 112307 - 0
 Intr 112556 112480 - 0
 Intr 112906 112856 - 0
 Init 113220 112995 - 0
 45 >4895176 /7663
 len = 2007 nex = 5
 Term 111441 111237 - 0
 50 Intr 112230 111955 - 0
 Intr 112394 112307 - 0
 Intr 112556 112480 - 0
 Init 112906 112856 - 0
 55 >4895176 /40538
 len = 2556 nex = 8
 Term 42176 41886 - 0
 60 Intr 42395 42282 - 0

1575

	Intr	42663	42480	-	0
	Intr	42988	42871	-	0
	Intr	43261	43133	-	0
	Intr	43455	43404	-	0
5	Intr	43792	43642	-	0
	Init	44441	44284	-	0
	>4895176 /109069				
10	len =	1787	nex =	2	
	Init	51902	51959	+	0
	Term	52945	53244	+	0
15	>4895176 /18810				
	len =	829	nex =	1	
20	Sngl	54124	53296	-	0
	>4895176 /11827				
	len =	862	nex =	1	
25	Sngl	54143	53282	-	0
	>4895176 /2445				
30	len =	2175	nex =	11	
	Term	65959	65856	-	0
	Intr	66105	66059	-	0
	Intr	66298	66203	-	0
	Intr	66481	66382	-	0
35	Intr	66642	66582	-	0
	Intr	66796	66725	-	0
	Intr	66979	66899	-	0
	Intr	67227	67159	-	0
	Intr	67424	67338	-	0
40	Intr	67812	67723	-	0
	Init	68030	67896	-	0
	>4895176 /912				
45	len =	2890	nex =	12	
	Term	65760	65465	-	0
	Intr	65959	65853	-	0
	Intr	66105	66059	-	0
50	Intr	66298	66203	-	0
	Intr	66481	66382	-	0
	Intr	66642	66582	-	0
	Intr	66796	66725	-	0
	Intr	66979	66899	-	0
55	Intr	67227	67159	-	0
	Intr	67424	67338	-	0
	Intr	67812	67723	-	0
	Init	68345	67896	-	0
60	>4895176 /10987				

1576

len = 914 nex = 1
 5 Sngl 75026 75939 + 0
 >4895176 /34976
 len = 986 nex = 1
 10 Sngl 84464 83479 - 0
 >4895213 /17709
 len = 215 nex = 1
 15 Sngl 12033 11819 - 0
 >4895213 /19452
 20 len = 1181 nex = 4
 Term 12235 11861 - 0
 Intr 12448 12322 - 0
 Intr 12634 12559 - 0
 25 Init 13041 12797 - 0
 >4895213 /168
 len = 1128 nex = 2
 30 Term 21760 21460 - 0
 Init 22587 22368 - 0
 >4895213 /35168
 35 len = 2411 nex = 6
 Init 69033 69478 + 0
 Intr 69674 69795 + 0
 40 Intr 69886 69976 + 0
 Intr 70067 70187 + 0
 Intr 70283 70356 + 0
 Term 70828 71443 + 0
 45 >4895233 /19714
 len = 1933 nex = 7
 Term 14019 13908 - 0
 50 Intr 14335 14258 - 0
 Intr 14918 14868 - 0
 Intr 15157 14999 - 0
 Intr 15354 15239 - 0
 Intr 15541 15460 - 0
 55 Init 15691 15621 - 0
 >4895233 /6457
 len = 2356 nex = 8
 60

1577

5	Term	14019	13655	-	0
	Intr	14335	14258	-	0
	Intr	14918	14868	-	0
	Intr	15157	14999	-	0
	Intr	15354	15239	-	0
	Intr	15541	15460	-	0
	Intr	15691	15621	-	0
Init	16010	15800	-	0	
10	>4895233 /19091				
	len =	1316	nex =	0	
15	>4895233 /2018				
	len =	944	nex =	0	
	>4895233 /34688				
20	len =	1882	nex =	4	
25	Term	57501	57086	-	0
	Intr	57848	57580	-	0
	Intr	58122	58008	-	0
	Init	58967	58722	-	0
30	>4914356 /19995				
	len =	1154	nex =	2	
	Term	25058	24628	-	0
	Init	25781	25678	-	0
35	>4914383 /41120				
	len =	1665	nex =	4	
40	Init	69870	70209	+	0
	Intr	70434	70564	+	0
	Intr	70654	70875	+	0
	Term	70961	71525	+	0
45	>4914383 /22376				
	len =	1122	nex =	5	
	Init	93930	94034	+	0
	Intr	94116	94290	+	0
50	Intr	94386	94491	+	0
	Intr	94572	94661	+	0
	Term	94741	95051	+	0
	>4914399 /36778				
55	len =	2278	nex =	8	
	Term	17584	17205	-	0
60	Intr	17738	17692	-	0
	Intr	18153	18069	-	0
	Intr	18331	18253	-	0
	>4914399 /36778				

1578

	Intr	18578	18425	-	0
	Intr	18790	18711	-	0
	Intr	19120	18889	-	0
	Init	19482	19368	-	0
5	>4914399 /97049				
	len =	537	nex =	1	
10	Sngl	61982	61446	-	0
	>4914400 /20435				
	len =	1768	nex =	7	
15	Term	20696	20292	-	0
	Intr	20940	20830	-	0
	Intr	21127	21041	-	0
	Intr	21289	21216	-	0
20	Intr	21473	21380	-	0
	Intr	21619	21574	-	0
	Init	22059	21711	-	0
	>4914400 /17491				
25	len =	491	nex =	1	
	Sngl	25638	25148	-	0
30	>4914400 /32381				
	len =	674	nex =	1	
	Sngl	30579	31252	+	0
35	>4914400 /124761				
	len =	709	nex =	1	
40	Sngl	52209	51501	-	0
	>4914400 /117732				
	len =	1754	nex =	5	
45	Term	85506	85180	-	0
	Intr	85754	85603	-	0
	Intr	86180	85944	-	0
	Intr	86678	86458	-	0
50	Init	86933	86763	-	0
	>4914400 /42211				
	len =	630	nex =	1	
55	Sngl	90161	89532	-	0
	>4914422 /30671				
60	len =	735	nex =	2	

2025 RELEASE UNDER E.O. 14176

1579

	Term	11060	10810	-	0
	Init	11544	11267	-	0
5	>4914422	/41723			
	len =	523	nex =	1	
10	Sngl	16294	16816	+	0
	>4914422	/15801			
	len =	1274	nex =	4	
15	Init	40051	40158	+	0
	Intr	40255	40329	+	0
	Intr	40525	40659	+	0
	Term	41043	41301	+	0
20	>4914422	/21466			
	len =	2006	nex =	3	
25	Init	46972	47324	+	0
	Intr	47772	47952	+	0
	Term	48182	48359	+	0
	>4914422	/28549			
30	len =	1090	nex =	3	
	Init	50748	50839	+	0
	Intr	51200	51370	+	0
	Term	51631	51834	+	0
35	>4914422	/147351			
	len =	1099	nex =	3	
40	Init	50774	50839	+	0
	Intr	51200	51370	+	0
	Term	51631	51859	+	0
	>4914422	/15569			
45	len =	268	nex =	1	
	Sngl	56936	56669	-	0
50	>4914422	/152675			
	len =	463	nex =	1	
	Sngl	57037	56575	-	0
55	>4914422	/32703			
	len =	1041	nex =	3	
60	Term	57059	56669	-	0

1580

	Intr	57259	57112	-	0
	Init	57709	57543	-	0
5	>4914422	/97001			
	len =	1020	nex =	1	
	Sngl	57709	56690	-	0
10	>4914422	/13439			
	len =	1019	nex =	2	
	Term	57259	56691	-	0
15	Init	57709	57543	-	0
	>4914422	/27704			
20	len =	989	nex =	2	
	Term	57259	57112	-	0
	Init	57709	57543	-	0
25	>4914422	/104060			
	len =	1030	nex =	3	
	Term	57059	56690	-	0
	Intr	57259	57112	-	0
30	Init	57710	57543	-	0
	>4914422	/32548			
35	len =	1050	nex =	2	
	Term	57259	56665	-	0
	Init	57714	57543	-	0
40	>4914422	/4740			
	len =	1222	nex =	2	
	Init	69466	69678	+	0
	Term	69803	69921	+	0
45	>4914422	/14629			
	len =	730	nex =	2	
50	Init	71747	71925	+	0
	Term	72026	72475	+	0
	>4914422	/34593			
55	len =	1780	nex =	5	
	Init	73912	74057	+	0
	Intr	74215	74328	+	0
	Intr	74412	74925	+	0
60	Intr	75022	75175	+	0

				1581	
	Term	75262	75691	+	0
	>4914449	/33307			
5	len =	1602	nex =	3	
	Term	9666	9504	-	0
	Intr	10587	10510	-	0
	Init	10901	10672	-	0
10	>4914449	/2508			
	len =	2620	nex =	6	
15	Term	32362	32080	-	0
	Intr	32797	32572	-	0
	Intr	33006	32910	-	0
	Intr	33341	33085	-	0
	Intr	33603	33434	-	0
20	Init	34699	34297	-	0
	>4914454	/20902			
	len =	1779	nex =	4	
25	Init	14139	14543	+	0
	Intr	14915	15176	+	0
	Intr	15255	15322	+	0
	Term	15419	15917	+	0
30	>4914454	/7447			
	len =	1873	nex =	2	
35	Term	16739	15950	-	0
	Init	17822	17623	-	0
	>4914454	/19084			
40	len =	339	nex =	1	
	Sngl	18789	19127	+	0
	>4914454	/118265			
45	len =	391	nex =	1	
	Sngl	33814	33429	-	0
50	>4914454	/11319			
	len =	2254	nex =	7	
	Init	4901	4989	+	0
55	Intr	5232	5572	+	0
	Intr	5666	5761	+	0
	Intr	5846	5923	+	0
	Intr	6011	6187	+	0
	Intr	6602	6670	+	0
60	Term	6755	7154	+	0

>4928204 /40175

	len =	2530	nex =	10	
5	Term	16431	15960	-	0
	Intr	16690	16520	-	0
	Intr	16930	16820	-	0
	Intr	17119	17018	-	0
10	Intr	17277	17215	-	0
	Intr	17440	17369	-	0
	Intr	17638	17537	-	0
	Intr	17812	17711	-	0
	Intr	17983	17911	-	0
15	Init	18119	18067	-	0

>4938473 /95855

	len =	940	nex =	2	
20	Init	34311	34428	+	0
	Term	34806	35250	+	0

>4938473 /21878

25	len =	2202	nex =	5	
	Term	75630	75293	-	0
	Intr	75942	75702	-	0
30	Intr	76292	76068	-	0
	Intr	76487	76363	-	0
	Init	77494	77121	-	0

>4938473 /2891

35	len =	1229	nex =	1	
	Sngl	9654	10033	+	0

40 >4938473 /125503

	len =	234	nex =	1	
	Sngl	9689	9922	+	0

45 >4938493 /21702

	len =	4165	nex =	13	
50	Init	13805	13896	+	0
	Intr	14185	14254	+	0
	Intr	14586	14698	+	0
	Intr	14905	15001	+	0
	Intr	15177	15256	+	0
55	Intr	15440	15634	+	0
	Intr	15720	15803	+	0
	Intr	15895	15955	+	0
	Intr	16076	16136	+	0
	Intr	16239	16332	+	0
60	Intr	16404	16476	+	0

2009-09-25 10:00:00

1583

	Intr	16614	16696	+	0
	Term	16814	17269	+	0
	>4938493 /14713				
5	len =	2612	nex =	6	
	Term	37708	37490	-	0
	Intr	37918	37806	-	0
10	Intr	38169	38036	-	0
	Intr	38398	38290	-	0
	Intr	38544	38485	-	0
	Init	38752	38670	-	0
15	>4938493 /26276				
	len =	310	nex =	1	
	Sngl	80685	80993	+	0
20	>4972043 /33435				
	len =	796	nex =	1	
25	Sngl	16528	17323	+	0
	>4972043 /22677				
	len =	674	nex =	1	
30	Sngl	18856	19529	+	0
	>4972043 /92148				
35	len =	970	nex =	4	
	Init	25133	25274	+	0
	Intr	25355	25571	+	0
	Intr	25661	25801	+	0
40	Term	25910	26100	+	0
	>4972043 /2545				
	len =	2009	nex =	3	
45	Init	2771	3001	+	0
	Intr	3642	3942	+	0
	Term	4073	4779	+	0
50	>4972043 /41319				
	len =	1930	nex =	3	
	Term	35309	35007	-	0
55	Intr	35674	35613	-	0
	Init	36929	36399	-	0
	>4972043 /35742				
60	len =	1947	nex =	2	

2025 RELEASE UNDER E.O. 14176

1584

	Init	49	578	+	0
	Term	771	1995	+	0
5	>4972043	/38287			
	len =	1172	nex =	1	
10	Sngl	62463	63634	+	0
	>4972043	/7471			
	len =	1810	nex =	6	
15	Init	92440	92483	+	0
	Intr	92601	92697	+	0
	Intr	92948	93047	+	0
	Intr	93126	93211	+	0
	Intr	93299	93390	+	0
20	Term	93561	93675	+	0
	>4972043	/4398			
	len =	1797	nex =	5	
25	Init	92470	92697	+	0
	Intr	92948	93047	+	0
	Intr	93126	93211	+	0
	Intr	93299	93390	+	0
30	Term	93561	93675	+	0
	>4972043	/93806			
	len =	1701	nex =	5	
35	Init	92571	92697	+	0
	Intr	92948	93047	+	0
	Intr	93126	93211	+	0
	Intr	93299	93390	+	0
40	Term	93561	93675	+	0
	>4972043	/119346			
	len =	647	nex =	0	
45	>4972065	/38344			
	len =	1428	nex =	3	
50	Term	31960	31284	-	0
	Intr	32312	32044	-	0
	Init	32711	32422	-	0
	>4972065	/34069			
55	len =	1213	nex =	3	
	Init	47383	47581	+	0
	Intr	47679	47884	+	0
60	Term	48236	48595	+	0

1585

```

>4972065      /10069
    len =      434    nex =      1
5      Sngl  48305  48738    +      0
>4972065      /36741
10     len =      1705   nex =      4
      Init  49180  49398    +      0
      Intr  49482  50010    +      0
      Intr  50089  50447    +      0
15     Term  50541  50884    +      0
>4972065      /101714
20     len =      294    nex =      1
      Sngl  52352  52059    -      0
>4972065      /35207
25     len =      1280   nex =      4
      Term  51432  51078    -      0
      Intr  51652  51532    -      0
      Intr  51983  51743    -      0
30     Init  52357  52061    -      0
>4972065      /126440
      len =      391    nex =      1
35     Sngl  59542  59932    +      0
>4972077      /12759
40     len =      2470   nex =      2
      Term  17826  17297    -      0
      Init  18688  17900    -      0
>4972077      /25532
45     len =      2549   nex =      2
      Term  17826  17288    -      0
      Init  18688  17900    -      0
50     >4972077      /116606
      len =      490    nex =      1
55     Sngl  58084  58567    +      0
>4972087      /38343
60     len =      817    nex =      1

```

1586

	Sngl	23707	24523	+	0
	>4972087	/28506			
5	len =	1074	nex =	3	
	Init	28873	29020	+	0
	Intr	29372	29517	+	0
	Term	29781	29946	+	0
10	>4972087	/30708			
	len =	941	nex =	1	
15	Sngl	30856	29916	-	0
	>4972087	/38500			
	len =	2072	nex =	7	
20	Init	33934	34065	+	0
	Intr	34547	34684	+	0
	Intr	34962	34999	+	0
	Intr	35096	35146	+	0
25	Intr	35230	35397	+	0
	Intr	35468	35580	+	0
	Term	35674	36005	+	0
	>4972087	/115901			
30	len =	1524	nex =	2	
	Term	8097	8047	-	0
	Init	9570	9200	-	0
35	>4996901	/2496			
	len =	1366	nex =	3	
40	Init	15599	15705	+	0
	Intr	15871	15982	+	0
	Term	16118	16964	+	0
	>4996901	/35390			
45	len =	2432	nex =	5	
	Term	872	613	-	0
	Intr	1483	1364	-	0
50	Intr	1822	1718	-	0
	Intr	2438	2370	-	0
	Init	3044	2539	-	0
	>4996901	/15624			
55	len =	1960	nex =	7	
	Init	55525	55641	+	0
	Intr	55761	55942	+	0
60	Intr	56052	56159	+	0

00940-9256500

1587

Intr	56321	56410	+	0
Intr	56669	56753	+	0
Intr	57080	57174	+	0
Term	57264	57484	+	0

5

>4996901 /11543

```
len = 2590    nex = 10
```

10	Init	78466	78573	+	0
	Intr	78960	79043	+	0
	Intr	79136	79248	+	0
	Intr	79343	79556	+	0
	Intr	79691	79764	+	0
15	Intr	79859	79912	+	0
	Intr	80008	80115	+	0
	Intr	80204	80262	+	0
	Intr	80494	80704	+	0
	Term	80794	81047	+	0

20

>4996901 /97320

```
len = 639    nex = 2
```

25	Term	81166	80888	-	0
	Init	81526	81238	-	0

>4996901 /36526

30	len =	1570	nex =	4	
	Init	81684	81730	+	0
	Intr	81908	82053	+	0
	Intr	82294	82409	+	0
35	Term	82501	83245	+	0

>4996901 /37444

```

40      len =      1696    nex =          3
      Init  83424  83855      +      0
      Intr  84133  84410      +      0
      Term  84602  85119      +      0

```

45 >4996901 /30264

```
len = 1299    nex = 2
```

	Term	91084	90551	-	0
50	Init	91849	91557	-	0

>4996902 /27688

```

55      len =      879    nex =      2
      Term 61399    61074    -      0
      Init 61655    61591    -      0

```

60 >4996902 /142223

1588

	len =	285	nex =	1	
	Sngl	83376	83092	-	0
5	>4996902	/11977			
	len =	2719	nex =	5	
10	Term	84065	83542	-	0
	Intr	84858	84288	-	0
	Intr	85264	84965	-	0
	Intr	85618	85341	-	0
	Init	86260	85915	-	0
15	>4996903	/21835			
	len =	2193	nex =	4	
20	Term	60521	59989	-	0
	Intr	60800	60607	-	0
	Intr	61272	60875	-	0
	Init	62181	61788	-	0
25	>5002514	/22413			
	len =	3101	nex =	12	
30	Term	22954	22566	-	0
	Intr	23196	23047	-	0
	Intr	23432	23292	-	0
	Intr	23601	23514	-	0
	Intr	23740	23685	-	0
	Intr	23895	23820	-	0
35	Intr	24045	23972	-	0
	Intr	24285	24127	-	0
	Intr	24518	24375	-	0
	Intr	24733	24601	-	0
	Intr	24951	24845	-	0
	Init	25666	25475	-	0
40	>5002514	/24475			
	len =	1016	nex =	2	
45	Term	34112	33856	-	0
	Init	34871	34579	-	0
	>5002514	/15906			
50	len =	4330	nex =	13	
	Term	47968	47811	-	0
	Intr	48454	48334	-	0
55	Intr	48638	48589	-	0
	Intr	48846	48733	-	0
	Intr	49060	48980	-	0
	Intr	49228	49147	-	0
	Intr	49372	49318	-	0
	Intr	49764	49534	-	0
60	Intr	50625	49920	-	0

1589

	Intr	51051	50964	-	0
	Intr	51239	51133	-	0
	Intr	51447	51329	-	0
	Init	52137	51799	-	0
5	>5002514 /15314				
	len =	1710	nex =	4	
10	Init	69964	70199	+	0
	Intr	70466	70639	+	0
	Intr	70727	70894	+	0
	Term	71131	71673	+	0
15	>5019262 /93927				
	len =	983	nex =	2	
	Term	59395	58861	-	0
20	Init	59843	59533	-	0
	>5019264 /110927				
	len =	415	nex =	1	
25	Sngl	4369	4766	+	0
	>5019264 /122052				
30	len =	370	nex =	1	
	Sngl	4404	4766	+	0
	>5019264 /23133				
35	len =	743	nex =	1	
	Sngl	98229	97508	-	0
40	>5019265 /2686				
	len =	1280	nex =	5	
	Term	83	1	-	0
45	Intr	243	167	-	0
	Intr	604	337	-	0
	Intr	923	718	-	0
	Init	1280	1010	-	0
50	>5041959 /117665				
	len =	778	nex =	1	
	Sngl	29026	29803	+	0
55	>5041960 /39877				
	len =	1241	nex =	1	
60	Sngl	33119	34355	+	0

1590

>5041962 /32146

5	len =	2478	nex =	7	
	Init	27173	27393	+	0
	Intr	27737	27765	+	0
	Intr	28119	28208	+	0
	Intr	28281	28520	+	0
10	Intr	28603	28800	+	0
	Intr	28954	29072	+	0
	Term	29160	29650	+	0

>5041962 /37936

15	len =	1330	nex =	4	
	Init	36213	36536	+	0
	Intr	36817	36933	+	0
20	Intr	37029	37116	+	0
	Term	37380	37537	+	0

>5041962 /28093

25	len =	1450	nex =	4	
	Init	36215	36536	+	0
	Intr	36817	36933	+	0
	Intr	37029	37116	+	0
30	Term	37380	37657	+	0

>5041962 /109138

35	len =	692	nex =	2	
	Init	36222	36536	+	0
	Term	36817	36901	+	0

>5041968 /92527

40	len =	1055	nex =	2	
	Term	39971	39857	-	0
45	Init	40467	40315	-	0

>5051726 /5364

	len =	1619	nex =	3	
50	Init	106365	106639	+	0
	Intr	106745	106850	+	0
	Term	107554	107983	+	0

>5051726 /19178

55	len =	1333	nex =	5	
	Init	18384	18450	+	0
	Intr	18595	18622	+	0
60	Intr	19014	19098	+	0

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which encodes an amino acid sequence exhibiting at least 40% sequence identity to an amino acid sequence encoded by

- (a) a nucleotide sequence described in Table 1 or a fragment thereof; or
(b) a complement of a nucleotide sequence shown in Table 1 or a fragment thereof.

2. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to

- (a) a nucleotide sequence shown in Table 1 or a fragment thereof; or
(b) a complement of a nucleotide sequence described in Table 1 or a fragment thereof.

3. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to a gene comprising

- (a) a nucleotide sequence shown in Table 1 or a fragment thereof; or
(b) a complement of a nucleotide sequence described in Table 1 or a fragment thereof.

4. An isolated nucleic acid molecule which is the reverse of the isolated nucleotide sequence according to claim 1, such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of said isolated nucleotide sequence according to claim 1.

5. An isolated nucleic acid molecule comprising a nucleic acid capable of hybridizing to a nucleic acid having a sequence selected from the group consisting of:

- (a) a nucleotide sequence which is shown in Table 1; and
(b) a nucleotide sequence which is complementary to a nucleotide sequence shown in Table 1;

under conditions that permit formation of a nucleic acid duplex at a temperature from about 40°C and 48°C below the melting temperature of the nucleic acid duplex.

6. The nucleic acid molecule according to claim 1, wherein said nucleic acid comprises an open reading frame.

7. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid is capable of functioning as a promoter, a 3' end termination sequence, an untranslated region (UTR), or as a regulatory sequence.

8. The isolated nucleic acid molecule of claim 7, wherein said nucleic acid is a promoter and comprises a sequence selected from the group consisting of a TATA box sequence, a CAAT box sequence, a motif of GCAATCG or any transcription-factor binding sequence, and any combination thereof.

9. The isolated nucleic acid molecule of claim 7, wherein the nucleic acid sequence is a regulatory sequence which is capable of promoting seed-specific expression, embryo-specific expression, ovule-specific expression, tapetum-specific expression or root-specific expression of a sequence or any combination thereof.

10. A vector construct comprising a nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is heterologous to any element in said vector construct.

11. A vector construct comprising:

- (a) a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation; and
- (b) a second nucleic acid having the sequence of the isolated nucleic acid molecule according to claim 1;

wherein said first and second nucleic acids are operably linked and wherein said second nucleic acid is heterologous to any element in said vector construct.

12. The vector construct according to claim 11, wherein said first nucleic acid is native to said second nucleic acid.

13. The vector construct according to claim 11, wherein said first nucleic acid is heterologous to said second nucleic acid.

14. A vector construct comprising:

- (c) a first nucleic acid having the sequence of the isolated nucleic acid molecule according to claim 7; and
- (d) a second nucleic acid;

wherein said first and second nucleic acids are operably linked and wherein said first nucleic acid is heterologous to any element in said vector construct.

15. The vector construct according to claim 14, wherein said first nucleic acid is native to said second nucleic acid.

16. The vector construct according to claim 14, wherein said first nucleic acid is heterologous to said second nucleic acid.

17. A host cell comprising an isolated nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is flanked by exogenous sequence.

18. A host cell comprising a vector construct of claim 10.

19. A host cell comprising a vector construct of claim 11.

20. A host cell comprising a vector construct of claim 12.

21. A host cell comprising a vector construct of claim 13.

22. A host cell comprising a vector construct of claim 14.

23. A host cell comprising a vector construct of claim 15.

24. A host cell comprising a vector construct of claim 16.

25. An isolated polypeptide comprising an amino acid sequence

(a) exhibiting at least 40% sequence identity of an amino acid sequence encoded by a sequence shown in Table 1 or a fragment thereof; and

5 (b) capable of exhibiting at least one of the biological activities of the polypeptide encoded by said nucleotide sequence shown in Table 1 or a fragment thereof.

26. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 75% sequence identity to an amino acid sequence encoded by a sequence shown in Table 1 or a fragment thereof.

27. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 85% sequence identity to an amino acid sequence encoded by a sequence shown in Table 1 or a fragment thereof.

28. The isolated polypeptide of claim 25, wherein said amino acid sequence exhibits at least 90% sequence identity to an amino acid sequence encoded by a sequence shown in Table 1 or a fragment thereof.

29. An antibody capable of binding the isolated polypeptide of claim 25.

30. A method of introducing an isolated nucleic acid into a host cell comprising:
(a) providing an isolated nucleic acid molecule according to claim 1; and
(b) contacting said isolated nucleic with said host cell under conditions that permit insertion of said nucleic acid into said host cell.

31. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 10.

32. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 11.

33. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 12.
34. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 13.
35. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 14.
36. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 15.
37. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 16.
38. A method of modulating transcription and/or translation of a nucleic acid in a host cell comprising:
 - (a) providing the host cell of claim 17; and
 - (b) culturing said host cell under conditions that permit transcription or translation.
39. A method for detecting a nucleic acid in a sample which comprises:
 - (a) providing an isolated nucleic acid molecule according to claim 1;
 - (b) contacting said isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of said isolated nucleic acid molecule with the sequence of DNA in said sample; and
 - (c) analyzing the result of said comparison.
40. The method according to claim 39, wherein said isolated nucleic acid molecule and said sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.
41. A plant or cell of a plant which comprises a nucleic acid molecule according to claim 1 which is exogenous to said plant or plant cell.
42. A plant or cell of a plant which comprises a nucleic acid molecule according to claim 1, wherein said nucleic acid molecule is heterologous to said plant or said cell of a plant.
43. A plant or cell of a plant which has been transformed with a nucleic acid molecule according to claim 1.
44. A plant or cell of a plant which comprises a vector construct according to claim 10.

45. A plant or cell of a plant which has been transformed with a vector construct according to claim 10.
46. A plant which has been regenerated from a plant cell according to claim 41.
47. A plant which has been regenerated from a plant cell according to claim 42.
48. A plant which has been regenerated from a plant cell according to claim 43.
49. A plant which has been regenerated from a plant cell according to claim 44.
50. A plant which has been regenerated from a plant cell according to claim 45.

46. A plant which has been regenerated from a plant cell according to claim 41.
47. A plant which has been regenerated from a plant cell according to claim 42.
48. A plant which has been regenerated from a plant cell according to claim 43.
49. A plant which has been regenerated from a plant cell according to claim 44.
50. A plant which has been regenerated from a plant cell according to claim 45.

47. A plant which has been regenerated from a plant cell according to claim 42.

48. A plant which has been regenerated from a plant cell according to claim 43.

49. A plant which has been regenerated from a plant cell according to claim 44.

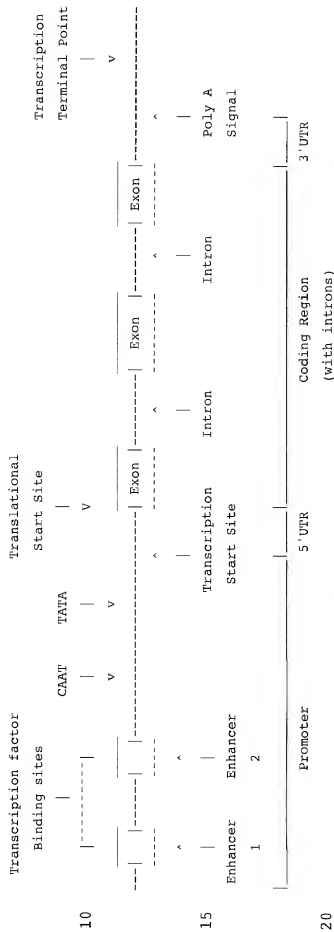
50. A plant which has been regenerated from a plant cell according to claim 45.

1597

SCHEMATIC I

SCHEMATIC OF A GENE

5



20

Sequences/motifs that influence specific DNA conformation, chromatin conformation, extent and position of base methylation and binding sites of proteins that control of these.

25

Gene

1598

ABSTRACT OF THE DISCLOSURE

The present invention provides DNA molecules that constitute fragments of the genome of a plant, and polypeptides encoded thereby. The DNA molecules are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome, in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait.

10

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Index**
 10. **Table of Contents**
 11. **Abstract**
 12. **Keywords**
 13. **Summary**
 14. **Notes**
 15. **References**
 16. **Appendix**
 17. **Index**
 18. **Table of Contents**
 19. **Abstract**
 20. **Keywords**
 21. **Summary**
 22. **Notes**
 23. **References**
 24. **Appendix**
 25. **Index**
 26. **Table of Contents**
 27. **Abstract**
 28. **Keywords**
 29. **Summary**
 30. **Notes**
 31. **References**
 32. **Appendix**
 33. **Index**
 34. **Table of Contents**
 35. **Abstract**
 36. **Keywords**
 37. **Summary**
 38. **Notes**
 39. **References**
 40. **Appendix**
 41. **Index**
 42. **Table of Contents**
 43. **Abstract**
 44. **Keywords**
 45. **Summary**
 46. **Notes**
 47. **References**
 48. **Appendix**
 49. **Index**
 50. **Table of Contents**
 51. **Abstract**
 52. **Keywords**
 53. **Summary**
 54. **Notes**
 55. **References**
 56. **Appendix**
 57. **Index**
 58. **Table of Contents**
 59. **Abstract**
 60. **Keywords**
 61. **Summary**
 62. **Notes**
 63. **References**
 64. **Appendix**
 65. **Index**
 66. **Table of Contents**
 67. **Abstract**
 68. **Keywords**
 69. **Summary**
 70. **Notes**
 71. **References**
 72. **Appendix**
 73. **Index**
 74. **Table of Contents**
 75. **Abstract**
 76. **Keywords**
 77. **Summary**
 78. **Notes**
 79. **References**
 80. **Appendix**
 81. **Index**
 82. **Table of Contents**
 83. **Abstract**
 84. **Keywords**
 85. **Summary**
 86. **Notes**
 87. **References**
 88. **Appendix**
 89. **Index**
 90. **Table of Contents**
 91. **Abstract**
 92. **Keywords**
 93. **Summary**
 94. **Notes**
 95. **References**
 96. **Appendix**
 97. **Index**
 98. **Table of Contents**
 99. **Abstract**
 100. **Keywords**
 101. **Summary**
 102. **Notes**
 103. **References**
 104. **Appendix**
 105. **Index**
 106. **Table of Contents**
 107. **Abstract**
 108. **Keywords**
 109. **Summary**
 110. **Notes**
 111. **References**
 112. **Appendix**
 113. **Index**
 114. **Table of Contents**
 115. **Abstract**
 116. **Keywords**
 117. **Summary**
 118. **Notes**
 119. **References**
 120. **Appendix**
 121. **Index**
 122. **Table of Contents**
 123. **Abstract**
 124. **Keywords**
 125. **Summary**
 126. **Notes**
 127. **References**
 128. **Appendix**
 129. **Index**
 130. **Table of Contents**
 131. **Abstract**
 132. **Keywords**
 133. **Summary**
 134. **Notes**
 135. **References**
 136. **Appendix**
 137. **Index**
 138. **Table of Contents**
 139. **Abstract**
 140. **Keywords**
 141. **Summary**
 142. **Notes**
 143. **References**
 144. **Appendix**
 145. **Index**
 146. **Table of Contents**
 147. **Abstract**
 148. **Keywords**
 149. **Summary**
 150. **Notes**
 151. **References**
 152. **Appendix**
 153. **Index**
 154. **Table of Contents**
 155. **Abstract**
 156. **Keywords**
 157. **Summary**
 158. **Notes**
 159. **References**
 160. **Appendix**
 161. **Index**
 162. **Table of Contents**
 163. **Abstract**
 164. **Keywords**
 165. **Summary**
 166. **Notes**
 167. **References**
 168. **Appendix**
 169. **Index**
 170. **Table of Contents**
 171. **Abstract**
 172. **Keywords**
 173. **Summary**
 174. **Notes**
 175. **References**
 176. **Appendix**
 177. **Index**
 178. **Table of Contents**
 179. **Abstract**
 180. **Keywords**
 181. **Summary**
 182. **Notes**
 183. **References**
 184. **Appendix**
 185. **Index**
 186. **Table of Contents**
 187. **Abstract**
 188. **Keywords**
 189. **Summary**
 190. **Notes**
 191. **References**
 192. **Appendix**
 193. **Index**
 194. **Table of Contents**
 195. **Abstract**
 196. **Keywords**
 197. **Summary**
 198. **Notes**
 199. **References**
 200. **Appendix**
 201. **Index**
 202. **Table of Contents**
 203. **Abstract**
 204. **Keywords**
 205. **Summary**
 206. **Notes**
 207. **References**
 208. **Appendix**
 209. **Index**
 210. **Table of Contents**
 211. **Abstract**
 212. **Keywords**
 213. **Summary**
 214. **Notes**
 215. **References**
 216. **Appendix**
 217. **Index**
 218. **Table of Contents**
 219. **Abstract**
 220. **Keywords**
 221. **Summary**
 222. **Notes**
 223. **References**
 224. **Appendix**
 225. **Index**
 226. **Table of Contents**
 227. **Abstract**
 228. **Keywords**
 229. **Summary**
 230. **Notes**
 231. **References**
 232. **Appendix**
 233. **Index**
 234. **Table of Contents**
 235. **Abstract**
 236. **Keywords**
 237. **Summary**
 238. **Notes**
 239. **References**
 240. **Appendix**
 241. **Index**
 242. **Table of Contents**
 243. **Abstract**
 244. **Keywords**
 245. **Summary**
 246. **Notes**
 247. **References**
 248. **Appendix**
 249. **Index**
 250. **Table of Contents**
 251. **Abstract**
 252. **Keywords**
 253. **Summary**
 254. **Notes**
 255. **References**
 256. **Appendix**
 257. **Index**

POWER OF ATTORNEY

CERES, INC.
3007 Malibu Canyon Road
Malibu, CA 90265

I, Richard Hamilton, Chief Financial Officer of CERES, INC. of 3007 Malibu Canyon Road, Malibu, California 90265, grant Power of Attorney and authority to empower the following attorneys to act on behalf of CERES, INC. for executing Verified Statements (Declarations) Claiming Small Entity Status to be submitted to the U.S. Patent and Trademark Office in connection with the filing of provisional or regular patent applications on behalf of CERES, INC.

Raymond C. Stewart (Reg. No. 21,066)
Joseph A. Kolasch (Reg. No. 22,463)
Leonard R. Svensson (Reg. No. 30,330)
Gerald M. Murphy, Jr. (Reg. No. 28,977)
Mark J. Nuell (Reg. No. 36,623)

This Power of Attorney is to remain in full force and effect until terminated by an official of CERES, INC.

By


Richard Hamilton

Date

9/24/98